

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 09:12:13 ; Search time 33.7584 Seconds
(without alignments)
9093.534 Million cell updates/sec

Title: US-09-691-220-3_COPY_5000_6000

Perfect score: 1001

Sequence: 1 tgatctccctccccccacc.....tgggtcaccctctggtg 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	52	5.2	7218	1	US-08-232-463-14
C 2	44.6	4.5	549	1	US-07-728-221B-12
C 3	44.6	4.5	549	3	US-07-728-220C-2
C 4	39.6	4.0	1894	3	US-07-728-220C-1
C 5	39	3.9	7218	1	US-08-232-463-14
C 6	38.4	3.8	2378	4	US-09-441-340-27
C 7	38	3.8	4032	4	US-09-068-101-5
C 8	37.4	3.7	5326	3	US-08-658-136-2
C 9	37.4	3.7	53577	3	US-08-658-136-1
C 10	36.6	3.7	2480	4	US-09-534-638-3
C 11	36.4	3.6	3070	1	US-08-428-733-3
C 12	36.4	3.6	111282	4	US-09-754-250-3
C 13	35.4	3.5	1969	4	US-09-098-628-1
C 14	34.8	3.5	320	4	US-09-165-264-7
C 15	34	3.4	320	4	US-09-165-264-13
C 16	34	3.4	1403	2	US-08-202-044-1
C 17	34	3.4	1403	4	US-08-751-344B-1
C 18	34	3.4	1614	1	US-08-583-673-1
C 19	33.8	3.4	75	4	US-09-461-697-151
C 20	33.8	3.4	320	4	US-09-165-264-14
C 21	33.8	3.4	517	1	US-08-480-784-33
C 22	33.8	3.4	517	1	US-08-483-553-33
C 23	33.8	3.4	517	1	US-08-487-002-33
C 24	33.8	3.4	517	1	US-08-483-554B-33
C 25	33.8	3.4	517	1	US-08-488-011B-33
C 26	33.8	3.4	517	4	US-08-850-727-33
C 27	33.8	3.4	517	5	PCT-US95-10202-33

Sequence 33, Appl
Sequence 33, Appl
Sequence 136, Appl
Sequence 57, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 7, Appl
Sequence 1, Appl
Sequence 323, Appl
Sequence 1, Appl
Sequence 18, Appl
Sequence 1, Appl
Sequence 129, Appl
Sequence 3, Appl
Sequence 318, Appl

C 28 33.8 3.4 517 5 PCT-US95-10203-33
C 29 33.8 3.4 517 5 PCT-US95-10220-33
C 30 33.8 3.4 542 4 US-09-461-697-136
C 31 33.8 3.4 1829 2 US-08-687-080-57
C 32 33.8 3.4 8438 1 US-07-945-283-1
C 33 33.6 3.4 8285 4 US-09-732-025-3
C 34 33.6 3.4 11827 4 US-09-739-455-3
C 35 33.6 3.4 44377 2 US-08-804-227C-7
C 36 33.6 3.4 44377 2 US-08-804-198-1
C 37 33.4 3.3 1166 4 US-09-072-596-323
C 38 33.4 3.3 1422 1 US-08-439-725A-1
C 39 33.4 3.3 1422 2 US-08-867-471-1
C 40 33.4 3.3 1422 3 US-08-705-245-18
C 41 33.4 3.3 1422 5 PCT-US96-06664-1
C 42 33.4 3.3 5590 4 US-09-050-159-129
C 43 33.4 3.3 17138 4 US-09-813-819-3
C 44 33.4 3.3 17138 4 US-09-920-048-3
C 45 33.2 3.3 712 4 US-09-149-476-318

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match

5.2%; Score 52; DB 1; Length 7218;

Best Local Similarity 4.3%; Pred. No. 2.7e-05;
Matches 16; Conservative 206; Mismatches 146; Indels 0; Gaps 0;

QY 193 GGGAGAGATAGTGGGAGAGAAACAGGACGAGGAGTGGTGGAGTGAAGAGCAGA 252
Db 1406 RRR 1347
QY 253 CGGGGTGGGAGGTCAGGAGAGATCTGCTGGGCTGGGATGGTGGGCAATCAACTGT 312
Db 1346 RRR 1287
QY 313 CCCATTGCTGAGGCTGCTTGGGAGGAGGAGGATGGGGGCCATAGCAGTCTGG 372
Db 1286 RRR 1227
QY 373 TCAGCCAGGCTGGCTGGAGTGGTGGCCAGGCACTACTAAGAGCCAGGAAGCCCTGC 432
Db 1226 RRR 1167
QY 433 CAAGGTGTTGGCTAGTTCCTGTCTATCAGCCGCTAGCAGCCGCCCTGCTGCAG 492
Db 1166 RRR 1107
QY 493 GTAAGGGGGAGGTGGTACATAGTACAGCCCTGCTGTTCCATGCTTCCCTCT 552
Db 1106 RRR 1047
QY 553 GTGCCCA 560
Db 1046 GCAGCCAA 1039

RESULT 2
US-07-728-221B-12/C
; Sequence 12, Application US/07728221B
; Patent No. 5340920
; GENERAL INFORMATION:
; APPLICANT: Matsuo, Hisayuki
; APPLICANT: Kangawa, Kenji
; APPLICANT: Miamoto, Naoto
; TITLE OF INVENTION: NOVEL HYDROLOGICALLY ACTIVE PORCINE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: ELEVENTH FLOOR, 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/728,221B
; FILING DATE: 19910712
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26,581
; REFERENCE/DOCKET NUMBER: WTS/9437/91817/KIK
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 671 4627 CUSH
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 549 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-07-728-221B-12

Query Match 4.5%; Score 44.6; DB 1; Length 549;
Best Local Similarity 49.5%; Pred. No. 0.0013;
Matches 145; Conservative 0; Mismatches 144; Indels 4; Gaps 1;

QY 89 TGGCTCACCCCTCTGCTCCCTCACCAGGCGCCCAACCGCCCGCCAGCAG 148
Db 306 TGTGCGCTTCTCTGACCGCCCGCCAGCTGGGCTGGCCACCTCTCCCTGGCG 247
QY 149 CTGTTCTCAGGCTCTCAGC---CTGCTGATTTCTGTTGCTGGGAGAGATGAG 204
Db 246 GAGTTTCGAGGACCTTCGCGCGCTCCGGGCTTGGCTTCGGAGGCGCGAGGAGGA 187
QY 205 GTGGGAGAGAAACAGGCGCAGGAGTGTGTTGGAGTGAAGAGCAGACGCGCGTGGGA 264
Db 186 GCGTAGCAGCAGAGAGCGCAGCAGCAGCTGGGAGAGGTGCATGGTGCCTGTGGGTGGA 127
QY 265 GGTGAGGAGAGATCTGCTGGGCTGGGATGGTGGGATCAACTGTCCCATTTGCTGCA 324
Db 126 GGGGCGCAGATCGGCGGAGAGAGTGCACGGGCGCGGCGGCGAGCTGCCAAGCGGCG 67
QY 325 GCGTGTCTTGGGCGAGGAGGAGGATGGGCGGCCATAGCAGTGTGGTGCAGC 377
Db 66 CGGAGCAGACTGCGCGGCTGCGGTGCGGCGCGAGTCCAGTGTCTGCGCGGC 14

RESULT 3
US-07-728-220C-2/C
; Sequence 2, Application US/07728220C
; Patent No. 6020168
; GENERAL INFORMATION:
; APPLICANT: MATSUO, HISAYUKI
; APPLICANT: KANGAWA, KENJI
; APPLICANT: TANAKA, SHOJI
; APPLICANT: FUCHIMURA, KAYOKO
; APPLICANT: TAWARAGI, YASUNORI
; TITLE OF INVENTION: PORCINE CNP GENE AND PRECURSOR PROTEIN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: ELEVENTH FLOOR, 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/728,220C
; FILING DATE: 19910712
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26,581
; REFERENCE/DOCKET NUMBER: WTS/9437/91816
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)861-3000
; TELEFAX: (202)822-0944
; TELEX: 671 4627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 549 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-728-220C-2

[illegible]

RESULT 5
US-08-232-463-14

; sequence 14, application 05/00432403
; Patent No. 5670367

; GENERAL INFORMATION:

APPLICANT: DORNER, F.
APPLICANT: SCHEITFLINGER, F.

;
;
; APPLICANT: SCHEIFLINGER, F.
;
; APPLICANT: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

```

; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:

```

```

;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
;

```

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA
COUNTRY: USA

COUNTRY: USA
ZIP: 22313-0299

COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

;
;
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
: APPLICATION NUMBER: 05/08/2332 163

```

APPLICATION NUMBER: US/08/232,463
FILING DATE:

CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; ADDITION NUMBER: US /07 /035 313

APPLICATION NUMBER: US/07/935,313
FILING DATE:

; FILING DATE: ;
 ; APPLICATION NUMBER: EP 91 114 300.6 ;

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.

NAME: BENI, STEPHEN A.
REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:
TELEPHONE. (703) 836-9300

TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109

TELEX: 899149

; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
```

; ENCODING: 740 base pairs
; TYPE: nucleic acid

STRANDEDNESS: single
MONOLOCV: linear

```

;
; TOPOLOGY: linear
;
; IMMEDIATE SOURCE:

```

```

; IMMEDIATE SOURCE:
; CLONE: PTZqpt-F1s
;

```

US-08-232-463-14

Overall Match 1 3 99: Score 30: PB 1: Length 7218:

Query Match . 3.9%; Score 39; DB 1; Length 7218;
Best Local Similarity 1.6%; Pred. No. 0.16;

best local similarity 1.00, freq. no. 0.10,

US-08-658-136-1/c
; Sequence 1, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D

```

1  /***** GENERAL INFORMATION *****/
2  /**** APPLICANT: Panula, Pertti A.J. ****/
3  /**** APPLICANT: Brandt, Annika ****/
4  /**** APPLICANT: Westerlund, Johanna ****/
5  /**** TITLE OF INVENTION: Promoter for ****/
6  /**** TITLE OF INVENTION: for therapy ****/
7  /**** FILE REFERENCE: 2530-104 ****/
8  /**** CURRENT APPLICATION NUMBER: US ****/
9  /**** CURRENT FILING DATE: 2000-03-22 ****/
10 /**** EARLIER APPLICATION NUMBER: 09 ****/
11 /**** EARLIER FILING DATE: 1999-08-01 ****/
12 /**** NUMBER OF SEQ ID NOS: 22 ****/
13 /**** SOFTWARE: PatentIn Ver. 2.1 ****/
14 /**** SEQ ID NO 3 ****/
15 /**** LENGTH: 2480 ****/

```

```

1  /***** GENERAL INFORMATION *****/
2  /**** APPLICANT: Panula, Pertti A.J. ****/
3  /**** APPLICANT: Brandt, Annika ****/
4  /**** APPLICANT: Westerlund, Johanna ****/
5  /**** TITLE OF INVENTION: Promoter for ****/
6  /**** TITLE OF INVENTION: for therapy ****/
7  /**** FILE REFERENCE: 2530-104 ****/
8  /**** CURRENT APPLICATION NUMBER: US ****/
9  /**** CURRENT FILING DATE: 2000-03-22 ****/
10 /**** EARLIER APPLICATION NUMBER: 09 ****/
11 /**** EARLIER FILING DATE: 1999-08-01 ****/
12 /**** NUMBER OF SEQ ID NOS: 22 ****/
13 /**** SOFTWARE: PatentIn Ver. 2.1 ****/
14 /**** SEQ ID NO 3 ****/
15 /**** LENGTH: 2480 ****/

```



```

RESULT 14
US-09-165-264-7
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamcoorthy, Thuraiyay
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

```

Search completed: March 30, 2003, 13:56:22
Job time : 174.758 secs

Search completed: March 30, 2003, 13:56:22
Job time : 174.758 secs

Db 104439 CCGCGCGSNNNDNNCCBGCGCGCGCGCGCCSNNNDNNCCBGCGCCBG 104499

QY 671 GGCGCGCGCTGGTGGGGTGTGTAAAGGGAGGACACCGGGACCAACCCCTCTTCC 730

Db 104499 CGCGCGCGCGSNNNDNNCCBGCGCGCGCGCGCGSNNNDNNCCBGCGCCBG 104558

QY 731 CGCGCGCGCGCTCTTCCACACGGCTTCGTGGCGCGCGCGCGCGCGCGCGCG 790

Db 104559 GGGCGCGCGCGSNNNDNNCCBGCGCGCGCGCGCGCGCGCGCGCGCGCG 104618

QY 791 GAGCTGGGAGCGCGAACTGTACAAGGGAGGACCGCGCGCGCTCTTCTTGGTTC 850

Db 104619 GCGCGCGCGSNNNDNNCCBGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104678

QY 851 CTTGCGAGCGCGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 910

Db 104679 GCGSNNNDNNCCBGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960

QY 911 CTGAGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 104738

Db 104739 CGCGCGCGCGCGCGSNNNDNNCCBGCGCGCGCGCGCGCGCGCGCGCGCGCG 104788

RESULT 10

ABA08208

ID ABA08208 standard; DNA; 13862 BP.

XX ABA08208;

AC ABA08208;

XX

DT 11-JAN-2002 (first entry)

DE Human ovarian and breast cancer associated polynucleotide SEQ ID NO 1003.

XX

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;

KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antiulcer;

KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;

KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;

XX neurological disease; infection; human; secreted protein; ds.

OS Homo sapiens.

XX

PN WO200155325-A2.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01345.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 28-JUN-2000; 2000US-0209467.

PR 30-JUN-2000; 2000US-0214886.

PR 07-JUL-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen -
is used in preventing, treating or ameliorating a medical condition -
Disclosure; SEQ ID NO 5477; 1297pp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.

SQ Sequence 13862 BP; 3188 A; 3560 C; 4465 G; 2649 T; 0 other;

Query Match 5.4%; Score 54; DB 22; Length 13862;

Best Local Similarity 47.0%; Pred. No. 0.013;

Matches 232; Conservative 0; Mismatches 260; Indels 2; Gaps 2;

XX Disclosure; Fig 1; 2lpp; English.

XX PS

XX CC The polypeptide may be used as a vaccine to provide

XX CC mammals, e.g. pigs, with resistance against Aujeszky's

XX CC disease.

XX CC

XX SQ Sequence 4897 BP; 594 A; 2052 C; 1495 G; 756 T; 0 other;

Query Match 5.2%; Score 51.6; DB 11; Length 4897;

Best Local Similarity 55.6%; Pred. No. 0.036;

Matches 99; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 823 GACGCGCGCCCTCTTCGGTCTGTGTCGCCCTCGAGAGCCCTCTCTCCCTGTACTCGCC 882

DB 3900 GTCCCTCTCGCCTCGGTCCCTCTCCCTCGCTCCTCTCCCTCGCTCCCTCCCTCC 3959

QY 883 GTCCCTCTGTACTCTGTGTACTCTCACTGTGAGAGCCTTTCCTCTCTCTCTCTCT 942

DB 3960 GTCCCTCTGCCCCCGCTGCTCTGCTCCCTCGCTCTGCTCCCTCGCTCTGCTCTGCC 4019

QY 943 CTCCTCCCTCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1000

DB 4020 CTCCTGCTCTCTCCCTCGCTCTCTCCCTCGCTCTCTCCCTCGCTCTCTCCCTCGCTCTCT 4077

RESULT 14

ABQ52496/C

ID ABQ52496 standard; DNA; 600 BP.

XX AC ABQ52496;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 39087.

XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX KW drug; side effect; cancer; central nervous system; cardiovascular;

XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;

XX KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP10074.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PR 05-SEP-2000; 2000DE-104543.

XX XX

PA (EPIG-) EPIGENOMICS AG.

XX XX

XX Olek A, Pispembrock C, Berlin K, Guetig D;

XX WIPI; 2002-371829/40.

XX XX

PT Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA

XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.

```

CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
CC
XX
SQ Sequence 600 BP; 26 A; 51 C; 478 G; 44 T; 1 other;

      Query Match          5.1%; Score 51.4; DB 24; Length 600;
      Best Local Similarity 48.5%; Pred. No. 0.028;
      Matches 142; Conservative 0; Mismatches 151; Indels 0; Gaps 0

Qy 708 CACCGGGACCAACCCCTTTCCCGCGCCCAACACCTCTCTCCACCAAGCTTGTGCTCGGC 767
Db 339 CCCCCCGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 280

Qy 768 CAGGACTGACCAACCTTGGGGAGACTGGGAGCGCGGAACCTGCTACAGGGGAGACGC 827
Db 279 CCCCCCAACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 220

Qy 828 CGCGCCCTTCCGCTTGTCTCCCTCGCAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCT 887
Db 219 CCCCCCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 160

Qy 888 TCTGTACTCTGTGTACTCTCTATCTATCTGGAGCCCTTCCCGCTTCTCTCTCTCTCTCT 947
Db 159 GCCCCCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 100

Qy 948 CCGCTTCCAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1000
Db 99 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 47

RESULT 15
ABQ52497
ID ABQ52497 standard; DNA; 600 BP.
XX
XX AC ABQ52497;
XX
XX DT 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 39088.
XX
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; ds.
XX
XX OS Homo sapiens.
XX
XX FN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP10074.
XX
XX PR 01-SEP-2000; 2000DE-1043826.
XX
XX PR 05-SEP-2000; 2000DE-1044543.
XX
XX PA (EPIG-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX DR WPI; 2002-371829/40.
XX
XX PT Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of

```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: March 30, 2003, 09:12:13 ; Search time 33.7247 Seconds
(without alignments)
9093.534 Million cell updates/sec

Title: US-09-691-220-3_COPY_1_1000
Perfect score: 1000
Sequence: 1 gtcctgggtagcatgtaca.....acatgcgcctctccctctc 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	72.6	7.3	7218	1	US-08-232-463-14	Sequence 14, Appl
2	59.2	5.9	53526	3	US-08-658-136-2	Sequence 2, Appl
3	59.2	5.9	53577	3	US-08-658-136-1	Sequence 1, Appl
4	51.6	5.2	4897	6	5196516-7	Patent No. 5196516
C 5	49.6	5.0	320	4	US-09-165-264-7	Sequence 7, Appl
C 6	49.6	5.0	320	4	US-09-165-264-14	Sequence 14, Appl
C 7	49.2	4.9	318	4	US-09-165-264-12	Sequence 12, Appl
C 8	48.6	4.9	319	4	US-09-165-264-8	Sequence 8, Appl
C 9	48.6	4.9	3978	3	US-08-726-214-1	Sequence 1, Appl
C 10	48.4	4.8	152331	3	US-09-128-155-16	Sequence 16, Appl
C 11	48	4.8	320	4	US-09-165-264-11	Sequence 11, Appl
C 12	47.8	4.8	320	4	US-09-165-264-13	Sequence 13, Appl
C 13	44	4.4	320	4	US-09-165-264-13	Sequence 13, Appl
C 14	43.8	4.4	319	4	US-09-165-264-8	Sequence 8, Appl
C 15	43.2	4.3	32207	2	US-08-770-379-20	Sequence 20, Appl
C 16	43.2	4.3	32207	4	US-08-757-669A-20	Sequence 20, Appl
C 17	43.2	4.3	32207	4	US-09-230-371A-20	Sequence 20, Appl
C 18	42.8	4.3	152331	3	US-09-128-155-16	Sequence 16, Appl
C 19	42.6	4.3	3978	3	US-08-726-214-1	Sequence 1, Appl
C 20	42	4.2	80245	4	US-09-078-294-4	Sequence 4, Appl
C 21	42	4.2	80595	4	US-09-078-294-3	Sequence 3, Appl
C 22	41.8	4.2	320	4	US-09-165-264-7	Sequence 7, Appl
C 23	41.4	4.1	320	4	US-09-165-264-11	Sequence 11, Appl
C 24	41.4	4.1	320	4	US-09-165-264-14	Sequence 14, Appl
C 25	41.4	4.1	950	4	US-08-931-789A-6	Sequence 6, Appl
C 26	41.4	4.1	950	4	US-09-062-451-6	Sequence 6, Appl
C 27	41.4	4.1	950	4	US-09-598-326-6	Sequence 6, Appl

28	41.2	4.1	1192	5	PCT-US93-06251-51	Sequence 51, Appl
29	41.2	4.1	1320	3	US-08-989-251-28	Sequence 28, Appl
30	41.2	4.1	1320	3	US-09-340-250-28	Sequence 28, Appl
31	41.2	4.1	1320	4	US-09-528-108-28	Sequence 28, Appl
C 32	41.2	4.1	4496	4	US-08-765-907A-6	Sequence 6, Appl
C 33	41	4.1	1500	4	US-09-593-711A-10	Sequence 10, Appl
C 34	40.8	4.1	289	4	US-09-007-005-17	Sequence 17, Appl
C 35	40.8	4.1	289	4	US-09-244-796-17	Sequence 17, Appl
C 36	40.4	4.0	801	2	US-08-770-379-16	Sequence 16, Appl
C 37	40.4	4.0	801	4	US-08-757-669A-16	Sequence 16, Appl
C 38	40.4	4.0	801	4	US-09-230-371A-16	Sequence 16, Appl
C 39	40.2	4.0	318	4	US-09-165-264-12	Sequence 12, Appl
40	40.2	4.0	1877	4	US-09-780-173A-10	Sequence 10, Appl
41	40	4.0	135	4	US-09-018-635-36	Sequence 36, Appl
42	40	4.0	801	2	US-08-770-379-16	Sequence 16, Appl
43	40	4.0	801	4	US-08-757-669A-16	Sequence 16, Appl
44	40	4.0	801	4	US-09-230-371A-16	Sequence 16, Appl
45	40	4.0	2907	2	US-09-018-628-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptz9pt-Fls
; US-08-232-463-14

Query Match 7.3%; Score 72.6; DB 1; Length 7218;


```

RESULT 4
5196516-7
; Patent No. 5196516
; APPLICANT: SCHREURS, CHRISTA S.;METTENLEITER, THOMAS C.
; SIMON, ARTHUR J.;LUKAS, NORMI.RZIHA, HANNS J.
; TITLE OF INVENTION: PSEUDORABIES VIRUS VACCINE
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/383,833
; FILING DATE: 21-JUL-1989
; SEQ ID NO:7:
; LENGTH: 4897
5196516-7

Query Match          5.2%; Score 51.6; DB 6; Length 4897;
Best Local Similarity 55.6%; Pred. No. 0.001;
Matches 99; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 823 GACGCCGCCCTCTTCGGTCTCCCTCCCTCGAGCCCTCTCTCTCCCTGTACTCGGC 882
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3900.GTCCCTCTCCCTCCGTCCTCCCTCCCTCCCTCTCCCTCGGTCCTCTCCCTCC 3959

QY 883 GTCCCTCTGTACTCTGTACTCTCTATCTCGAGCCCTTCCCTCTCTCTCTCTCT 942
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3960 GTCCCTCTCCCTCCCTCCCTCTCTCCCTCTCCCTCTCTCCCTCCCTCTCTCC 4019

QY 943 CTCCTCCCTCTCCAGGCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1000
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4020 CTCGGTCTCTCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4077

RESULT 5
US-09-165-264-7/c
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

Query Match          5.0%; Score 49.6; DB 4; Length 320;
Best Local Similarity 47.7%; Pred. No. 0.0013;
Matches 145; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 697 TAAGGGAGGACACCGGGACACCCCTCTTCCCGCCGCCACACCTCTCTCCACCGG 756
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 318 TAAGGGATGGTAGCTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 259

QY 757 CTTGCTCGGCCAGGGACTGACCAAACTTTGGGGAGCCTTGGAGCGGAACTGGTACAA 816
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 258 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 199

QY 817 GGGAGGAGCGCGCCCTCTTTCGCTCTTGTCCCTTCGAGCCCTCTCTCTCTCTGTA 876
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 198 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 139

QY 877 CTGGCGCTCTCTGTACTCTGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 936
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 138 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 79

QY 937 TCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 996
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Best Local Similarity 48.0%; Pred. No. 0.0017;
Matches 141; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 707 ACACGGGACACACCCCTTCTCCCGCCGCCACCACTCTCTCCACCAAGGCTTCGCTCGG 766
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 309 ACACACCAACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 250
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 767 CCAGGACACTGACAAACCTTGGGGAGCCTTGGAGCCGGAATGTTACAAGGGAGGACG 826
    || || || || || || || || || || || || || || || || || || || ||
Db 249 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 190
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 827 CCGCCGCTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 886
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 189 CCGCCGCTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 130
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 887 CTCTGTACTCTGTGTACTCTCATCTGGAGGCTTTCGCCCTTCTGCTTCCTCTCTCC 946
    || || || || || || || || || || || || || || || || || || || ||
Db 129 CCGCCGCTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 70
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 947 TCCGCTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1000
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69 CCGCCGCTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 16

RESULT 8
US-09-165-264-8/C
; Sequence 8, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiyath
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ. ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-8

Query Match 4.9%; Score 48.6; DB 4; Length 319;
Best Local Similarity 47.8%; Pred. No. 0.0024;
Matches 141; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 706 GACACGGGACACCCCGCTTCTCCCGCCGCCACCACTCTCTCCACCAAGGCTTCGCTCG 765
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 302 GACCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 243
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 766 GCACGGGACTGACAAACCTTGGGGAGCCTTGGAGCCGGAATGTTACAAGGGAGGAC 825
    || || || || || || || || || || || || || || || || || || || ||
Db 242 CCGCCGCTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 183
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 826 GCGCGGCTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 885
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 182 CCGCGGCTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 123
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 886 CTCTGTACTCTGTGTACTCTCATCTGGAGGCTTTCGCCCTTCTGCTTCCTCTCTCC 945
    || || || || || || || || || || || || || || || || || || || ||
Db 122 CCGCCGCTTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 70
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 946 CTCCGCTTTCGAGGCTGCGCCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1000
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 62 CCGCGGCTTTCGAGGCTGCGCCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 8
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

```

```

; GENERAL INFORMATION:
; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gilman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,214
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,498
; FILING DATE: 04-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSD:450
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3978 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-726-214-1

Query Match 4.9% Score 48.6; DB 3; Length 3978;
Best Local Similarity 58.08; Pred. No. 0.0053;
Matches 105; Conservative 0; Mismatches 74; Indels 2; Gaps

QY 452 ACTTGGCCCCCGGAGACCCGGCCCTACGCCCTCTCTGCC- -GCCGCTCTCCGCGTCTCCGGG 509
Db 13 ACTGGCGCGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 72
QY 510 GGAGTGGCCCGGTTCCGCGCGGCGAGGGGCTGGCGGCGAGCCCGCGGCGGCGTGGC 569
Db 73 CGAGGCGCGCGCGCGCGCTGCGTGGCGTGGGATGGCGGGGCGCGCGCGCGCGCGCGCG 132
QY 570 GAGCGGTGATGTACCGGCGAGCGGTGGTGGTCTACTCGAGGTTGAGCGCGCGCGAGCG 629
Db 133 GCGGAGGCGGAGCGCGCGCGCGAGTCTGTGGGGGCGCGCGAGCGGCGCGCGCGCGCG 192
QY 630 G 630
Db 193 G 193

RESULT 10
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650

```

```
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; FEATURE:
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match          4.8%; Score 48.4; DB 3; Length 152331;
Best Local Similarity 51.7%; Pred. No. 0.018;
Matches 134; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 456 GGGCCCGCCGACCCGCGCTAGCCTCTCGCGCGCTCTCCGCTCTCCGGGGAGGT 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22139 GGGCGGGCGGGGGGGGGGGTGGCTGCGGGGGGGGGGGGGGGGGGGGGGGGG 22080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 516 GGGCGGGTTCGCGCGGAGGGGGCTGGCGGGAGCGCCCGCGGGGGGGTGGCGAGCGG 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22079 GGTGGGTGGCGGGGGTGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGG 22020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 576 GTGATGTACGGCGACGCTGGTGGGTGCTACCTCGGAGGTGAGCGCGCCGAGCGAGTTC 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22019 GTGGGGG---GCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 21963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 636 ACAGAGATTACGCCCATGTATGAGCAATGAGCGCGCGCTGGTGGGGGGGTGTGTG 695
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21962 CCGCGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 21903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 696 TTAAGGGGAGGACCGGG 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21902 GGGGGGGGGGGGGGGGG 21884

RESULT 11
US-09-165-264-11/c
; Sequence 11, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-11

Query Match          4.8%; Score 48; DB 4; Length 320;
Best Local Similarity 47.1%; Pred. No. 0.0034;
Matches 147; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 689 GTGTGTGTTAAGGGGAGACACCGGACCCCTCTTCCCGCCCGCCACCACTCTCTC 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 319 GTCTTCTTAAATGTCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 749 CACCACGGCTTCGCTCGCGCCAGGGGACTGACCAAACTTGGGGGAGCTGGGAGCGGAAC 808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 259 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 809 TGGTACAAGGGGAGACCGCCCGCCCTCTTCGCTGCTTCCCTCGCAGACCCCTCTCTC 868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; FEATURE:
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match          4.8%; Score 48.4; DB 3; Length 152331;
Best Local Similarity 51.7%; Pred. No. 0.018;
Matches 134; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 456 GGGCCCGCCGACCCGCGCTAGCCTCTCGCGCGCTCTCCGCTCTCCGGGGAGGT 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22139 GGGCGGGCGGGGGGGGGGGTGGCTGCGGGGGGGGGGGGGGGGGGGGGGGGG 22080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 516 GGGCGGGTTCGCGCGGAGGGGGCTGGCGGGAGCGCCCGCGGGGGGGTGGCGAGCGG 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22079 GGTGGGTGGCGGGGGTGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGG 22020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 576 GTGATGTACGGCGACGCTGGTGGGTGCTACCTCGGAGGTGAGCGCGCCGAGCGAGTTC 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22019 GTGGGGG---GCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 21963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 636 ACAGAGATTACGCCCATGTATGAGCAATGAGCGCGCGCTGGTGGGGGGGTGTGTG 695
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21962 CCGCGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 21903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 696 TTAAGGGGAGGACCGGG 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21902 GGGGGGGGGGGGGGGGG 21884

RESULT 12
US-09-165-264-13/c
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13

Query Match          4.8%; Score 47.8; DB 4; Length 320;
Best Local Similarity 47.8%; Pred. No. 0.0038;
Matches 139; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 710 CCGGACACCCCGCTTTCGCGCCCGCCACACCTCTCCACACGAGGTTCGCTCGGCA 769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 306 CCACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 770 GGACTGACCAAACTTGGGGGAGCTGGGAGCGGAACTGGTACAAAGGGGAGGACGCC 829
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 246 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 830 GCGCGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 186 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 890 TGTACTCTGTGTACTCTCTCATCTGAGGACCTTCCCGCTTCTCTCTCTCTCTCT 949
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 950 CCGTTCGCGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 66 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16

RESULT 13
US-09-165-264-13
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
```


Db 303 DSGSESSWSTGSSNGHGGGNGHKPGCEKPGNEARGSGESGICQFRQGVSSNM 362
Qy 624 CCAGCGGAGTTACGAGAGATTACGCCGCAATTCGATTAGG 663
Db 363 EISKEGNRLGSGDNYRGQSSWGGGDAVGWNTVNS 402

RESULT 7

US-09-954-456-292
; Sequence 292, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Candi
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 292
; LENGTH: 43058
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-292

Query Match 4.8%; Score 48.4; DB 10; Length 43058;
Best Local Similarity 52.4%; Pred. No. 0.00056;
Matches 130; Conservative 0; Mismatches 116; Indels 2; Gaps 1;
Qy 498 CGCGTCTCCGGGGAGGTGCGCGCGGCGAGG--GGGCTGGCGGGCGAGCCCC 555
Db 25640 CGCGGGCGGGCGGGTTCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 25699
Qy 556 CGCGGGGGGTGCGGAGCGGGTGATGTCACGGGACGCGGTGGGTGCTACTCGGAGGTG 615
Db 25700 GGGCCCGGGCGGGCGGGCGGGGAGGGGCTGGCGGGCGGGCGGGCGGGG 25759
Qy 616 AGCGCGCGCCAGCGAGTTTCAGCGAGAGTTCAGCCGCAATTCGATTAGGCAATGAGCCCC 675
Db 25760 CGGGCGGGCGGGCGGGTTCGCGGGCGGGGTCGCGGGCGGGCGGGCGGGCGG 25819
Qy 676 GGCGTGGGTGGGGTGTGTGTTAAGGGAGGACACCGGGGACACCCCGCTTCCCGGCC 735
Db 25820 GGGCGGGGTGGGTTCGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 25879
Qy 736 CCACCAACC 743
Db 25880 CCGCCCCC 25887

RESULT 8

US-09-954-456-529

; Sequence 529, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 529
; LENGTH: 43058
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-529

Query Match 4.8%; Score 48.4; DB 10; Length 43058;
Best Local Similarity 52.4%; Pred. No. 0.00056;
Matches 130; Conservative 0; Mismatches 116; Indels 2; Gaps 1;
Qy 498 CGCGTCTCCGGGGAGGTGCGCGCGGCGAGG--GGGCTGGCGGGCGAGCCCC 555
Db 25640 CGCGGGCGGGCGGGTTCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 25699
Qy 556 CGCGGGGGGTGCGGAGCGGGTGATGTCACGGGACGCGGTGGGTGCTACTCGGAGGTG 615
Db 25700 GGGCCCGGGCGGGCGGGCGGGGAGGGGCTGGCGGGCGGGCGGGCGGGG 25759
Qy 616 AGCGCGCGCCAGCGAGTTTCAGCGAGAGTTCAGCCGCAATTCGATTAGGCAATGAGCCCC 675
Db 25760 CGGGCGGGCGGGCGGGTTCGCGGGCGGGGTCGCGGGCGGGCGGGCGGGCGG 25819
Qy 676 GGCGTGGGTGGGGTGTGTGTTAAGGGAGGACACCGGGGACACCCCGCTTCCCGGCC 735
Db 25820 GGGCGGGGTGGGTTCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGG 25879
Qy 736 CCACCAACC 743
Db 25880 CCGCCCCC 25887

RESULT 9

US-09-880-107-3950
; Sequence 3950, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO

```
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3950
; LENGTH: 43058
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 Z84721
US-09-880-107-3950

Query Match          4.8%; Score 48.4; DB 10; Length 43058;
Best Local Similarity 52.4%; Pred. No. 0.00056;
Matches 130; Conservative 0; Mismatches 116; Indels 2; Gaps 1;

QY 498 CCGCTCTCCGGGAGGTGCGCGGTTCCGCGGCGAGG--GGGCTCGCGGCGAGCCCC 555
Db 25640 CCGCGGCGCGCGGCGGCGGTCGCGGGGCGGCGGCGGCGGCGGCGGCGGCGCG 25699

QY 556 GCGGGCGGCGCTGCGGAGCGGTGATCTCACGGGCAGCGGTGGTGTCTACTCGGAGGTG 615
Db 25700 GGGCCCGCGGCGGCGGCGGCGGCGGAGGCGCTGGCGGGCGGCGGCGGCGGCGGG 25759

QY 616 AGGCGCGCGCAGCGAGTTCACGAGAGTTCACCGCATTTAGGCAAAATCAGGCCCC 675
Db 25760 CCGGCGGCGCGCGGCGGCGGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 25819

QY 676 GCGCTGGTGGGCGTGTGTTAAGGGAGGACACCGGAGACACCCCTCTTCCCGGCC 735
Db 25820 GGGCGGGGTTGGGTCGCGGGGCGGCGGCGGCGGCTAGGCGCGCGCGCGCGCGCC 25879

QY 736 CCACCACC 743
Db 25880 CCGGCCCC 25887

RESULT 10
US-10-095-407-16/c
; Sequence 16, Application US/10095407
; Patent No. US20020164330A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/10/095,407
; CURRENT FILING DATE: 2002-03-11
; PRIOR APPLICATION NUMBER: US 60/091,650
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/054,646
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-10-095-407-16

Query Match          4.8%; Score 48.4; DB 9; Length 152331;
Best Local Similarity 51.7%; Pred. No. 0.00068;
Matches 134; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 456 GCGCCCGCGGACCGCGCCCTACGCTCTCGCGCGGCTCTCCGCGGAGGT 515
```

```
Db 22139 GGGCGGCGGCGGCGGCGGCTCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGG 22080
QY 516 GGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 575
Db 22079 GGTGCGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 22020
QY 576 GTGATGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 635
Db 22019 GTGGGGG---GCGCGGGGGGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21963
QY 636 AGCGAGAGTTACGCCGATTCATTAGGCAAAATGAGGCCCGGCGGCGGCGGCGGCGG 695
Db 21962 CGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21903
QY 696 TTAAGGGGAGACACCGGG 714
Db 21902 GGGGGGGGGGGGGGGGG 21884

RESULT 11
US-09-771-208-20
; Sequence 20, Application US/09771208
; Patent No. US20020155564A1
; GENERAL INFORMATION:
; APPLICANT: MEDRANO, JUAN
; APPLICANT: BRADFORD, ERIC
; APPLICANT: HORVAT, SIMON
; TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE
; FILE REFERENCE: 407T-923710US
; CURRENT APPLICATION NUMBER: US/09/771,208
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 08/999,477
; PRIOR FILING DATE: 1997-12-29
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 659158
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (123459)..(123478)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (602466)..(602485)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (546998)..(547017)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (494715)..(494814)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (390986)..(391005)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (346860)..(346823)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (317174)..(317193)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (280353)..(280373)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (271829)..(271848)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (183872)..(183891)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (170625)..(170645)
```


;; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;; TITLE OF INVENTION: ATHEROSCLEROSIS
;; FILE REFERENCE: 10797-004001
;; CURRENT APPLICATION NUMBER: US/10/023,523
;; CURRENT FILING DATE: 2001-12-17
;; PRIOR APPLICATION NUMBER: US/09/616,289
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 09/517,849
;; PRIOR FILING DATE: 2000-03-02
;; PRIOR APPLICATION NUMBER: US 08/979,608
;; PRIOR FILING DATE: 1997-11-26
;; PRIOR APPLICATION NUMBER: US 60/031,930
;; PRIOR FILING DATE: 1996-11-27
;; PRIOR APPLICATION NUMBER: US 60/048,547
;; PRIOR FILING DATE: 1997-06-03
;; NUMBER OF SEQ ID NOS: 53
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 48
;; LENGTH: 2561
;; TYPE: DNA
;; ORGANISM: Oryctolagus cuniculus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (246)...(1895)
US-10-023-523-48

Query Match 4.5%; Score 45; DB 12; Length 2561;
Best Local Similarity 54.7%; Pred. No. 0.0032;
Matches 110; Conservative 0; Mismatches 90; Indels 1; Gaps 1;
Qy 505 CGGGGGAGGTGCGCGGTTCGGCGGCGAGGGGCTGGCGGCGAGCCCGCGCGCGG 564
Db 854 CGGGGGGCGG 795
Qy 565 CTGGCGAGCGGTGATGTCACGGGAGCGGTGGTGGTCACTCGGAGGTGAGCGCCCG 624
Db 794 CGGCGCGCGCGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 736
Qy 625 CAGCGGAGTTCAGCGAGGTTCAGCGCGCATTCATTAGCAATAGCGCCCGCGCTGGCT 684
Db 735 GCGAGGGGCGCGTGTGGCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGG 676
Qy 685 GGGGTGTGTGTTAAGGGGAG 705
Db 675 CGGCGCGCGCGCGAGCGCG 655

RESULT 15
US-10-032-393-47
; Sequence 47, Application US/10032393
; Publication No. US20030027286A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Wall, Daniel
; APPLICANT: Gross, Molly
; TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
; FILE REFERENCE: ELITRA.010A
; CURRENT APPLICATION NUMBER: US/10/032,393
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/259,434
; PRIOR FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 09/948,993
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 12733
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

;; OTHER INFORMATION: Vector pEPEF14
US-10-032-393-47
Query Match 4.5%; Score 45; DB 9; Length 12733;
Best Local Similarity 48.0%; Pred. No. 0.0042;
Matches 129; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
Qy 507 GGGGGAGGTGCGCGGTTCGGCGGCGAGGGGCTGGCGGCGAGCCCGCGCGCGGCT 566
Db 5144 GGG 5203
Qy 567 GCGAGCGGTGATGTCACGGGCGAGCGGTGGTGGTCACTCGGAGGTGAGCGCCCGCA 626
Db 5204 GGG 5263
Qy 627 GCGAGTTCAGCGAGAGTTTCAGCGCGCATTCATTAGCAATAGCGCCCGCGCTGGTGG 686
Db 5264 GGG 5323
Qy 687 GGGTGTGTGTTAAGGGAGGACACCGGGAGCACCCCGCTTCCCGCGCCCGACCTCC 746
Db 5324 GGG 5383
Qy 747 TCCACCACGGCTTCGCTCGGCCAGGGACT 775
Db 5384 GCTCCTTCCGGTGGCGCGGGGCGATGACT 5412

Search completed: March 30, 2003, 16:45:13
Job time : 956.893 secs

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="6294603"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: xhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT	205 a	297 c	264 g	163 t
ORIGIN				
Query Match	65.6%; Score 656.2; DB 14; Length 929;			
Best Local Similarity	98.7%; Pred. No. 7.5e-141;			
Matches 672; Conservative	0; Mismatches 8; Indels 1; Gaps 1;			
QY	1	GTCCCTGGGTAGCATGATTCATCCATCCCTCCCTTTTATATATATATAGGGGTAAATAGATAC	60	
Db	696	GTCCCTGGGTAGCATGATTCATTCAT-CCTTCTTTAGAGAGTGGGGTAAATAGATAC	638	
QY	61	CCCTCTCCAGGGGTATCCCTCTTCTAGGACCTACCCAGCTAGGCTTTCTTTCCA	120	
Db	637	CCCTCTCCAGGGGTATCCCTCTTCTAGGACCTACCCAGCTAGGCTTTCTTTCCA	578	
QY	121	GTGAACGTGTCATCCAGGGGTTCTAGGATGAAGTAGTCCACTGGAAGCCACAGCTCT	180	
Db	577	GTGAACGTGTCATCCAGGGGTTCTAGGATGAAGTAGTCCACTGGAAGCCACAGCTCT	518	
QY	181	TCCTTTATCTCTCCAGAGCTGACAGTGCACAGGGGGCGGTACTGGTTCCTCCAGCTAG	240	
Db	517	TCCTTTATCTCTCCAGAGCTGACAGTGCACAGGGGGCGGTACTGGTTCCTCCAGCTAG	458	
QY	241	GAGACACCTTGGCGGGGTTTCTCGCGGAGACGACGAGCGGTGGGAGGAGGGGCC	300	
Db	457	GAGACACCTTGGCGGGGTTTCTCGCGGAGACGACGAGCGGTGGGAGGAGGGGCC	398	
QY	301	CCTCTGCGGTGTTTGTGCAACAGACCCGCGCTCCGCGTCCGGTTCGGCGGGCGGA	360	
Db	397	CCTCTGCGGTGTTTGTGCAACAGACCCGCGCTCCGCGTCCGGTTCGGCGGGCGGA	338	
QY	361	GTACACATGATGTCACACATGACACAGCCGGTGTCTCATTCGACACAGCGTCCG	420	
Db	337	GTACACATGATGTCACACATGACACAGCCGGTGTCTCATTCGACACAGCGTCCG	278	
QY	421	AGCTGCACATGTCACACCCGGGTGCCAACACTTGGCCCGCGGACCGGCCCTACGC	480	
Db	277	AGCTGCACATGTCACACCCGGGTGCCAACACTTGGCCCGCGGACCGGCCCTACGC	218	
QY	481	CTCTGCGCGCGTCTCCGCGGTCTCCGGGGAGGTGGCCCGGTTCGGCGGGCAGGGGC	540	
Db	217	CTCTGCGCGCGTCTCCGCGGTCTCCGGGGAGGTGGCCCGGTTCGGCGGGCAGGGGC	158	
QY	541	TGGCGGGCAGCCCGCGGGCGGTGGCGCGGGGTGTGTACCGGGCAGCGGTGGGTG	600	
Db	157	TGGCGGGCAGCCCGCGGGCGGTGGCGCGGGGTGTGTACCGGGCAGCGGTGGGTG	98	
QY	601	GGTCACTCGAGGTGAGGCGCGCCAGGCGAGTTTACGAGAGGTTCAGCGCGCATTCATT	650	
Db	97	GGTCACTCGAGGTGAGGCGCGCCAGGCGAGTTTACGAGAGGTTCAGCGCGCATTCATT	38	
QY	661	AGGCAAAATGAGCGCGGCGCTG	681	
Db	37	AGGCAAAATGAGCGCGGCGCTG	17	
RESULT 2				
BM702028/c	602 bp mRNA linear EST 28-FEB-2002			
LOCUS	BM702028			

DEFINITION	UI-E-CQ1-aex-p-12-0-UI_r1 UI-E-CQ1 Homo sapiens cDNA clone			
ACCESSION	UI-E-CQ1-aex-p-12-0-UI 5', mRNA sequence.			
VERSION	BM702028			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 602)			
TITLE	Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery			
JOURNAL	Genome Res. 6 (9), 791-806 (1996)			
MEDLINE	97044477			
COMMENT	Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@blue.weeg.uiowa.edu Tissue Procurement: Dr. Gregg Hageman cDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). The following repetitive elements were found in this cDNA sequence: 136-189, >(CGGG)n#simple_repeat (matched complement) Seq primer: M13 Reverse.			
FEATURES	Location/Qualifiers			
source	1..602			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="UI-E-CQ1-aex-p-12-0-UI"			
	/clone_lib="UI-E-CQ1"			
	/tissue_type="optic nerve"			
	/dev_stage="adult"			
	/lab_host="DH10B (Life Technologies) (Tl phage resistant)"			
	/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CQ1 is a normalized cDNA library containing the following tissue(s): optic nerve. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCATTAAAGT. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."			
BASE COUNT	103 a	216 c	184 g	98 t
ORIGIN	1 others			
Query Match	59.0%; Score 590; DB 14; Length 602;			
Best Local Similarity	99.7%; Pred. No. 1.2e-125;			
Matches 601; Conservative	0; Mismatches 1; Indels 1; Gaps 1;			
QY	111	CTTTCTCCAGTGAACGTGCATCCCGAGGCTTCTAGGATGAAGTAGTCCACATGGAAGG	170	
Db	602	CTTTCTCCAGTGAACGTGCATCCCGAGGCTTCTAGGATGAAGTAGTCCACATGGAAGG	543	
QY	171	CACCACTCTCTCTTTATCTCTCCAGAGCTGGACATGCACACAGGGCCGCTACTGCTT	230	
Db	542	CACCACTCTCTCTCTTTATCTCTCCAGAGCTGGACATGCACACAGGGCCGCTACTGCTT	484	
QY	231	CCCCAGTAGGACACACTTGGGGGGGCTTGTCTCCCGGAAGCACGACGAGCTGGGG	290	

```
Db 483 CCCAGCTAGGAGACACCTTTGGCGGGCTTGTCTCCGGAAGACGACGAGCGTGGG 424
QY 291 AGAGGGCCCCCTGCTGCTGTTGTGCAACAGCACCAGCGCTGCGGCTCGGGTCC 350
Db 423 AGAGGGCCCCCTGCTGCTGTTGTGCAACAGCACCAGCGCTGCGGCTCGGGTCC 364
QY 351 GGGCGCGGAGTGCACATGATGTCACAGACAATGACACAAAGCCGGTGTCTCATTCGGAC 410
Db 363 GGGCGCGGAGTGCACATGATGTCACAGACAATGACACAAAGCCGGTGTCTCATTCGGAC 304
QY 411 ACAGGTCGCGAGTGCACAATGTCACACCCGGGTGCCAACACTTGGCCCGCGACCC 470
Db 303 ACAGGTCGCGAGTGCACAATGTCACACCCGGGTGCCAACACTTGGCCCGCGACCC 244
QY 471 GGCCTACGCTCTCTGCGCGGCTCTCCGGTCTCCGGGGAGGTGCGCGGTCGGCG 530
Db 243 GGCCTACGCTCTCTGCGCGGCTCTCCGGTCTCCGGGGAGGTGCGCGGTCGGCGG 184
QY 531 GGCAGGGGGTGGCGGCGAGCCCGCGGGCTGGCGAGCGGGTGTGATGTCACGGCA 590
Db 183 GGCAGGGGGTGGCGGCGAGCCCGCGGGCTGGCGAGCGGGTGTGATGTCACGGCA 124
QY 591 GCGGTGGGTGGTCACTCGGAGGTGAGCGCGCGGCGAGTTCAGCGAGTTCAGCC 650
Db 123 GCGGTGGGTGGTCACTCGGAGGTGAGCGCGCGGCGAGTTCAGCGAGTTCAGCC 64
QY 651 GCATTGCTATGGAATAGGCGCGGCTGGGGTGGGGTGTGTGTTAAGGGGAGGAC 710
Db 63 GCATTGCTATGGAATAGGCGCGGCTGGGGTGGGGTGTGTGTTAAGGGGAGGAC 4
QY 711 CGG 713
Db 3 CGG 1
```

```
RESULT 3
BM713229/c
LOCUS
DEFINITION
  BM713229 603 bp mRNA linear EST 28-FEB-2002
  UI-E-EJ0-ahn-h-15-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
  BM713229
  BM713229.1 GI:19026487
  EST.
SOURCE
  human.
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 603)
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
JOURNAL
  Genome Res. 6 (9), 791-806 (1996)
MEDLINE
  97044477
COMMENT
  Contact: Soares, MB
  Program for Rat Gene Discovery and Mapping
  University of Iowa
  451 Eckstein Medical Research Building Iowa City, IA 52242, USA
  Tel: 319 335 8250
  Fax: 319 335 9565
  Email: msoares@blue.weeg.uiowa.edu
  Tissue Procurement: Dr. Gregg Hageman
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com).
  The following repetitive elements were found in this cDNA
  sequence: 138-191, >(CGGG)n$imple_repeat (matched complement)
  Seq primer: M13 Reverse.
  Location/Qualifiers
    1..603
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
```

FEATURES
source

```
/clone="UI-E-EJ0-ahn-h-15-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJ0 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dr)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye-Institute (NEI)."
```

QY	Db	BASE COUNT	105 a	203 c	199 g	96 t
QY 113	Db 603	TTCTTCCAGTGAACGTCATCCCGAGGGCTTCTAGGATGAAGTAGTCCACCTGGAAGGCA	172			
QY 173	Db 543	CCAGCTCTTCCTTTATCTCTCCAGAGCTGGACAGTGCACACAGGCGCGGTACTGTTCC	232			
QY 233	Db 483	CCAGCTAGGAGACACACTTTGGCGGGGCTTTGCTCCGCGAAGCAGCAGAGCGTGGGAG	292			
QY 293	Db 423	GAGGGCCCCCTCTGCTGCTGTTGTGCCAACAGACCCGCGCTCGCGGTTCGGTCCGG	352			
QY 353	Db 473	GAGGGCCCCCTCTGCTGCTGTTGTGCCAACAGACCCGCGCTCGCGGTTCGGTCCGG	364			
QY 413	Db 473	AGCGTCCGAGTGCACATGTCACACCCGGGTGCCAACACTTGGCCCCCGCGACCCCGG	472			
QY 473	Db 243	CCCTACGCTCTCTGCGCGGCTCTCCGGGTGTCTCCGGGGAGGTGCGCGGTTCGGCGGG	532			
QY 533	Db 183	CAGGGGCTGGCGGCGAGCCCGCGGGCTGGCGAGCGGTGTGTCACGGGCGACG	592			
QY 593	Db 123	GGTGGGTGGGTCACTCGGAGGTGAGCGCGCGGAGGTTCAGCGAGGTTTCAGCCCG	652			
QY 653	Db 63	ATTGCATTAGGCAAAATAGGCCCGCGCTG 681				
QY 681	Db 63	ATTGCATTAGGCAAAATAGGCCCGCGCTG 35				

Query Match 56.7%; Score 567.4; DB 14; Length 603;
Best Local Similarity 99.8%; Pred. No. 1.9e-120;
Matches 568; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 4

```
BM726559/c
LOCUS       BM726559      591 bp      mRNA      linear      EST 01-MAR-2002
DEFINITION  UI-E-EJ0-aii-i-18-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
ACCESSION   BM726559
VERSION     BM726559.1 GI:19047892
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 591)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            The following repetitive elements were found in this cDNA
            sequence: 132-185, >(CGGG)n#Simple_repeat (matched complement)
            Seq primer: M13 Reverse.
            Location/Qualifiers
            1..591
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="UI-E-EJ0-aii-i-18-0-UI"
               /clone_lib="UI-E-EJ0"
               /tissue_type="fetal eyes, lens, eye anterior segment,
               optic nerve, retina, Retina Foveal and Macular, RPE and
               Choroid"
               /dev_stage="fetal and adult"
               /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
               /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
               modified polylinker; Site_1: EcoR I; Site_2: Not I;
               UI-E-EJ0 is a subtracted cDNA library constructed
               according to Bonaldo, Lennon and Soares, Genome Research,
               6:791-806, 1996. First strand cDNA synthesis was primed
               with an oligo-dT primer containing a Not I site. Double
               stranded cDNA was ligated to an EcoR I adaptor, digested
               with Not I, and cloned directionally into pT73-Pac
               vector. The oligonucleotide used to prime the synthesis of
               first-strand cDNA contains a library tag sequence that is
               located between the Not I site and the (dT)18 tail. The
               sequence tags for this library are: fetal eyes, AGAATCAAGA
               ; lens, CCAATAGCGA; eye anterior segment, AATGCCGCAT;
               optic nerve, CCAATAGTG; retina, CCGCG; Retina Foveal and
               Macular, GTCC; RPE and Choroid, ACCTA. This library was
               created for the program, Gene Discovery in the Visual
               System, supported by National Eye Institute (NEI)."
BASE COUNT  101 a 202 c 193 g 95 t
ORIGIN
Query Match      56.3%; Score 563; DB 14; Length 591;
Best Local Similarity 100.0%; Pred. No. 1.9e-119;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 119 CAGTGAACGTCATCCCGAGGGCTTCTAGGATGAAGTAGTCCATCGAAGCCACCACT 178
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 591 CAGTGAACGTCATCCCGAGGGCTTCTAGGATGAAGTAGTCCATCGAAGCCACCACT 532
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 179 CTTCCTTTTATCTCTCCAGACTGGACAGTGCACCGGGCGGGTACTGGTTCCCGAGCT 238
```

```
|||||
Db 531 CTTCCTTTTATCTCTCCAGACTGGACAGTGCACCGGGCGGGTACTGGTTCCCGAGCT 472
      |||||||
Qy 239 AGGAGACACCTTGGCGGGGCTTTCCTCCGCGAAGACAGCAGAGCGTGGGAGGAGGC 298
      |||||||
Db 471 AGGAGACACCTTGGCGGGGCTTTCCTCCGCGAAGACAGCAGAGCGTGGGAGGAGGC 412
      |||||||
Qy 299 CCCCTCTGCTGTGTTCGCAACAGCAGCCCGCTGCCGGTCCGGTTCGGGCGGCCG 358
      |||||||
Db 411 CCCCTCTGCTGTGTTCGCAACAGCAGCCCGCTGCCGGTTCGGGTTCCGGCGGCCG 352
      |||||||
Qy 359 GAGTCACACATGATGTACAGACAATGACACAAGCGGGTGTCTCATTCGACACACGTC 418
      |||||||
Db 351 GAGTCACACATGATGTACAGACAATGACACAAGCGGGTGTCTCATTCGACACACGTC 292
      |||||||
Qy 419 CGAGTCGACAAATGTCACACCGGGTGCACAAACACTTGGCCCGCGCACCGGCCCTAC 478
      |||||||
Db 291 CGAGTCGACAAATGTCACACCGGGTGCACAAACACTTGGCCCGCGCACCGGCCCTAC 232
      |||||||
Qy 479 GCCTCTCCGCGCTCTCCGCTCTCCGGGGAGGTGGCGGGTTCGGCGGGGAGGG 538
      |||||||
Db 231 GCCTCTCCGCGCTCTCCGCTCTCCGGGGAGGTGGCGGGTTCGGCGGGGAGGG 172
      |||||||
Qy 539 GCTGCGGGGAGCGCGCGGGCTGGCGAGCGGGTGTACACGGCAGCGGTGG 598
      |||||||
Db 171 GCTGCGGGGAGCGCGCGGGCTGGCGAGCGGGTGTACACGGCAGCGGTGG 112
      |||||||
Qy 599 TGGGTCACTCGGAGGTGAGCGCGCCCGAGGCTTTCAGCGAGAGTTCAGCCGATGCA 658
      |||||||
Db 111 TGGGTCACTCGGAGGTGAGCGCGCCCGAGGCTTTCAGCGAGAGTTCAGCCGATGCA 52
      |||||||
Qy 659 TTAGCAAAATGAGCGCGCGCTG 681
      |||||||
Db 51 TTAGCAAAATGAGCGCGCGCTG 29
      |||||||

RESULT 5
LOCUS       BM931567/c
DEFINITION  UI-E-EJ1-ajk-j-22-0-UI.r1 UI-E-EJ1 Homo sapiens cDNA clone
ACCESSION   BM931567
VERSION     BM931567.1 GI:19390740
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 579)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     9704477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            The following repetitive elements were found in this cDNA
            sequence: 114-167, >(CGGG)n#Simple_repeat (matched complement)
            Seq primer: M13 REVERSE.
            Location/Qualifiers
            1..579
               /organism="Homo sapiens"
```


Db 180 CCCTCTCTCCAGGGGTATCCCTCTTCTAGGACCTACCCAAAGCTAGGCGCTTCTTCCA 239
QY 121 GTGAAGCTGCATCCCGAGGCTTCTAGGATGAAGTAGTCCATGGAGGACACAGCTCT 180
Db 240 GTGAAGCTGCATCCCGAGGCTTCTAGGATGAAGTAGTCCATGGAGGACACAGCTCT 299
QY 181 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGGCGGCTACTGGTTCCCGCAGTAG 240
Db 300 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGGCGGCTACTGGTTCCCGCAGTAG 359
QY 241 GAGACACTTGGGGGGCTTGTGCGCGAAGCAGCAGAGCGTGGGAGGAGGCC 300
Db 360 GAGACACTTGGGGGGCTTGTGCGCGAAGCAGCAGAGCGTGGGAGGAGGCC 419
QY 301 CCTCTGCTGTGTTGTGCCAAGCAGACCCGCGCTGCCGGTGGGTTCCGGCGCGGA 360
Db 420 CCTCTGCTGTGTTGTGCCAAGCAGACCCGCGCTGCCGGTGGGTTCCGGCGCGGA 479
QY 361 GTCACATGATGTACAGACATGACACAAGCGGTGTCTCATTTCCGACACAGCGTCCG 420
Db 480 GTCACATGATGTACAGACATGACACAAGCGGTGTCTCATTTCCGACACAGCGTCCG 539
QY 421 AGCTGCACATGTACACCGGGTGCACAACACTTGGCGCGCGGACCGGCGCTTACGC 480
Db 540 AGCTGCACATGTACACCGGGTGCACAACACTTGGCGCGCGGACCGGCGCTTACGC 599
QY 481 CTCCTGCGCGCTCTCCGGGTCTCCGGGAGGTGCCGGTGGGTCGGCGCGGC 533
Db 600 CTCCTGCGCGCTCTCCGGGTCTCCGGGAGGTGCCGGTGGGTCGGCGCGGC 652

RESULT 7
BM684023
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

BM684023 627 bp mRNA linear EST 27-FEB-2002
UI-E-EJ1-ajk-j-22-0-UI.s1 UI-E-EJ1 Homo sapiens cDNA clone
UI-E-EJ1-ajk-j-22-0-UI 3', mRNA sequence.

BM684023
EST.
BM684023.1 GI:18993919
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 627)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Forward
POLYA=Yes.

Location/Qualifiers
1. 627
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ1-ajk-j-22-0-UI"
/clone_lib="UI-E-EJ1"
/tissue.type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and

FEATURES
Source

Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJ1 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCGCGAT;
optic nerve, CCATTAAAGT; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI).
TAG_LIB=UI-E-EJ1
TAG_TISSUE=RPE and Choroid
TAG_SEQ=ACCTA"

BASE COUNT 123 a 185 c 164 g 155 t
ORIGIN

Query Match 49.5%; Score 494.6; DB 14; Length 627;
Best Local Similarity 98.2%; Pred. No. 1.2e-103;
Matches 500; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GTCCTTGGGTAGCATGATTCATCCCTCTCTTTTATATATGCGGGTATAGATAC 60
Db 119 GTCCTTGGGTAGCATGATTCATCCCTCTTTTAGAGAGTGGGGTATAGATAC 178

* QY 61 CCCCTCTCCAGGGGTATCCCTCTTTTAGGGAGCTACCCAAAGCTAGGCTTTCTTCCA 120
Db 179 CCCCTCTCCAGGGGTATCCCTCTTTTAGGGAGCTACCCAAAGCTAGGCTTTCTTCCA 238

QY 121 GTGAACGTCATCCGAGGCTTCTTAGGATGAAGTAGTCCACTGGAAGGCACAGCTCT 180
Db 239 GTGAACGTCATCCGAGGCTTCTTAGGATGAAGTAGTCCACTGGAAGGCACAGCTCT 298

QY 181 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGCGCGCTACTGTTCCCGAGCTAG 240
Db 299 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGCGCGCTACTGTTCCCGAGCTAG 358

QY 241 GAGACACTTGGCGGGGCTTTGCTCGCGGAAGCAGCAGAGCGTGGGGAGGAGGCC 300
Db 359 GAGACACTTGGCGGGGCTTTGCTCGCGGAAGCAGCAGAGCGTGGGGAGGAGGCC 418

QY 301 CCTCTGCTGTGTTGTGCCAAGCAGCAGCGGTGCCGCTGGGTTCCGGCGCGCGGA 360
Db 419 CCTCTGCTGTGTTGTGCCAAGCAGCAGCGGTGCCGCTGGGTTCCGGCGCGCGGA 478

QY 361 GTCACATGATGTACAGCAATGACACAAGCGGTGTCTCATTTCCGACAGAGCTCG 420
Db 479 GTCACATGATGTACAGCAATGACACAAGCGGTGTCTCATTTCCGACAGAGCTCG 538

QY 421 AGCTGCACATGTACACCGGGTGCACAACACTTGGCGCGCGGACCGGCGCTTACGC 480
Db 539 AGCTGCACATGTACACCGGGTGCACAACACTTGGCGCGCGGACCGGCGCTTACGC 598

QY 481 CTCCTGCGCGCTCTCCGGGTCTCCGGGTCTCCGGG 509
Db 599 CTCCTGCGCGCTCTCCGGGTCTCCGGGTCTCCGGG 627

RESULT 8
BM675116
LOCUS
DEFINITION
ACCESSION

BM675116 631 bp mRNA linear EST 27-FEB-2002
UI-E-EJ0-ahn-h-16-0-UI.s1 UI-E-EJ0 Homo sapiens cDNA clone
UI-E-EJ0-ahn-h-16-0-UI 3', mRNA sequence.
BM675116

```

VERSION      BM6751116.1  GI:18985014
KEYWORDS     EST.
SOURCE       Homo sapiens
ORGANISM     Human.
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        1 (bases 1 to 631)
JOURNAL      Bonaldo, M.F., Lennon, G. and Soares, M.B.
MEDLINE      Normalization and subtraction: two approaches to facilitate gene
COMMENT      discovery
              Genome Res. 6 (9), 791-806 (1996)
              9704477
              Contact: Soares, MB
              Program for Rat Gene Discovery and Mapping
              University of Iowa
              451 Eckstein Medical Research Building Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: msoares@blue.weeg.uiowa.edu
              Tissue Procurement: Dr. Gregg Hageman
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Researchers may obtain clones from Research
              Genetics (www.resgen.com).
              Seq primer: M13 Forward
              Polrfa=yes.

FEATURES     Location/Qualifiers
              1..631
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="UI-E-EJ0-ahn-h-16-0-UI"
                /clone_lib="UI-E-EJ0"
                /tissue_type="fetal eyes, lens, eye anterior segment,
                optic nerve, retina, Retina Foveal and Macular, RPE and
                Choroid"
                /dev_stages="fetal and adult"
                /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
                modified polylinker; Site_1: EcoR I; Site_2: Not I;
                UI-E-EJ0 is a subcloned cDNA library constructed
                according to Bonaldo, Lennon and Soares, Genome Research,
                6:791-806, 1996. First strand cDNA synthesis was primed
                with an oligo-dT primer containing a Not I site. Double
                stranded cDNA was ligated to an EcoR I adaptor, digested
                with Not I, and cloned directionally into pT73-Pac
                vector. The oligonucleotide used to prime the synthesis of
                first-strand cDNA contains a library tag sequence that is
                located between the Not I site and the (d18)18 tail. The
                sequence tags for this library are: fetal eyes, AGAATCAGA
                ; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
                optic nerve, CCATTAAAGTG; retina, CCGCG; Retina Foveal and
                Macular, GTCC; RPE and Choroid, ACCTA. This library was
                created for the program, Gene Discovery in the Visual
                System, supported by National Eye Institute (NEI).
                TAG_L1B=UI-E-EJ0
                TAG_TISSUE=human fetal eyes
                TAG_SEQ=AGAATCAAGA"
BASE COUNT   124 a 184 c 168 g 155 t
ORIGIN
Query Match      48.5%; Score 485; DB 14; Length 631;
Best Local Similarity 97.9%; Pred. No. 1.9e-101;
Matches 502; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 1 GTCCTGGGTACATGATCCATCCCTTCCTTTATATATATGGGGTAATAGATAC 60
DB 120 GTCTTGGGTACATGATCCATCCCTTCCTTTAGAGAGTGGGGTAATAGATAC 179
QY 61 CCCCTCCCTCAGGGGTATCCCTCTTTCTAGGAGACTACCCAGCTAGGCCCTTCTTCCA 120
DB 180 CCCCTCCCTCAGGGGTATCCCTCTTTCTAGGAGACTACCCAGCTAGGCCCTTCTTCCA 239

```

```

QY 121 GTGAACGTCATCCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGACCAAGCTCT 180
DB 240 GTGAACGTCATCCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGACCAAGCTCT 299
QY 181 TCCTTTTATCTCTCCAGAGTGGACAGTGCACACAGGGCCGGTACTGTTCCCAAGCTAG 240
DB 300 TCCTTTTATCTCTCCAGAGTGGACAGTGCACACAGGGCCGGTACTGTTCCCAAGCTAG 359
QY 241 GAGACACCTTTGGGGGGGGCTTTGCTCCGCCGGAACAGCAGAGCGTGGGAGAGGGCCC 300
DB 360 GAGACACCTTTGGGGGGGGCTTTGCTCCGCCGGAACAGCAGAGCGTGGGAGAGGG-CC 418
QY 301 CCTCTGCTGCTGTTTGGCCAAACAGACCCGCGCTCGCGCTCGCGTTCGGGCGGCCGA 360
DB 419 CCCTGCTGCTGTTTGGCCAAACAGACCCGCGCTCGCGCTCGCGTTCGGGCGGCCGAA 478
QY 361 GTCACACATGATGTCACAGACAAATGACACAAAGCCGGTGTCTCATTTCCGACACAGCGTCCG 420
DB 479 GTCACACATGATGTCACAGACAAATGACACAAAGCCGGTGTCTCATTTCCGACACAGCGTCCG 538
QY 421 AGCTGCAATATGTCACACCCGGGTGCCAAACACTTGGCCCGCGAGACCCGGCCCTACGC 480
DB 539 AGCTGCAATATGTCACACCCGGGTGCCAAACACTTGGCCCGCGAGACCCGGCCCTACGC 598
QY 481 CTCTGCGCGCGCTCTCCGGGTCTCCGGGGGAG 513
DB 599 CTCTGCGCGCGCTCTCCGGGTCTCCGGGGGAG 631

RESULT 9
BI759786/c
LOCUS      BI759786
DEFINITION 603045668F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5185961 5',
            mRNA sequence.
ACCESSION  BI759786
VERSION     BI759786.1  GI:15751364
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 797)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cyabs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLML1464 row: g column: 18
            High quality sequence stop: 582.
            Location/Qualifiers
              1..797
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5185961"
                /clone_lib="NIH_MGC_116"
                /lab_host="DH10B"
                /note="Organ: pooled colon, kidney, stomach; Vector:
                PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                source anonymous pool of 3 colons, age 26 yo male, 49 yo
                female, 71 yo male colon; 46 yo male kidney, and pool of 2
                stomachs, 62 yo male and 70 yo female. Library is
                oligo-dT primed and directionally cloned (EcoRV site is
                destroyed upon cloning). Average insert size 1.4 kb,
                insert size range 1-3 kb. Library is normalized and
                enriched for full-length clones and was constructed by C.
                Gruber (Invitrogen). Research Genetics tracking code

```

023. Note: this is a NIH_MGC Library."

```
BASE COUNT      140 a  280 c  234 g  143 t
ORIGIN
Query Match      48.28; Score 482; DB 13; Length 797;
Best Local Similarity 92.5%; Pred. No. 1e-100;
Matches 618; Conservative 0; Mismatches 35; Indels 15; Gaps 10;
QY  29 CTTCTCTTTTATATGGGGTATAGGATACCC--CTCCCTCAGGGGTATCCCTC-- 84
DB  596 CTTGTTAGAGAGTGGGGGTATAGGATACCCGGTGTGAGGGGTATCGCTCGT 637
QY  85 TTTCTAGGACCTACCAAGCTAGCCCTTTCTTCAGTG----AAACGTGCATCCCGAGG 140
DB  636 GTGCTAGGACCTAGCCGAGCTAGGCGCTTCCGTTCCAGGTGAACACCTGCATCCGAGG 577
QY  141 G-CTTCTAGGTAGGT-AGTCCACTGGAAGGCACCAGCTTTCC-TTTTATCTTCAG 197
DB  576 GCCTTCTAGGATGAAGTAAGTCCACTGGAAGGCACCAGCTTCTCGGGTGTATCTCCAG 517
QY  198 AGCTGGACAGTGCACCAAGGCCCGGTA-CTGGTTCCTCCAGCTAGGAGACACTTGGCGG 256
DB  516 AGCTGGACAGTGCACCAAGGCCCGGTACTGTTCCAGCTAGGAGACACTTGGCGG 457
QY  257 GG-CTTTGCTGCGCGGAAGCAGCAGAGCGTGGGAGAGGGCCCCCTC-TGCCCTGTG-T 313
DB  456 GGCCTTAGCTCGCGGGAAGCAGCAGAGCGTGGGAGAGGGCCCCCTCGTGCCTGTGT 397
QY  314 TTGTGCCAACAGCACCGCTGCGGTGCGGTTCGGGCGCGGCGGAGTACACATGATG 373
DB  396 TTGTGCCAACAGCACCGCTGCGGTGCGGTTCGGGCGCGGCGGAGTACACATGATG 337
QY  374 TCACAGACAATGACACAGCGGTGTCTATTCCGACACAGCGTCCGAGCTGCACAAATG 433
DB  336 TCACAGACAATGACACAGCGGTGTCTATTCCGACACAGCGTCCGAGCTGCACAAATG 277
QY  434 CACACCGGGTGCCAAACACTTGGCCCGCGCGAGCGCCCTACGCCCTCTGCGCGCGC 493
DB  276 CACACCGGGTGCCAAACACTTGGCCCGCGCGAGCGCCCTACGCCCTCTGCGCGCGC 217
QY  494 TCTCGCGTCTCCGGGAGTGGCCGGTTCGGCGCGGAGGGGCTGGCGGGGAGGCC 553
DB  216 TCTCGCGTCTCCGGGAGTGGCCGGTTCGGCGCGGAGGGGCTGGCGGGGAGGCC 157
QY  554 CCGCGGGGCGGCTGGCGAGCGGTGTCTACGGGACGCGTGGTGGGTCACTCGGAGG 613
DB  156 CCGCGGGGCGGCTGGCGAGCGGTGTCTACGGGACGCGTGGTGGGTCACTCGGAGG 97
QY  614 TGAGGCGCGCGAGCGGAGTTACGAGAGTTACGCCCATTTGCATTAGGCAAAATGAGGC 673
DB  96 TGAGGCGCGCGAGCGGAGTTACGAGAGTTACGCCCATTTGCATTAGGCAAAATGAGGC 37
QY  674 CCGGCCCTG 681
DB  36 CCGGCCCTG 29

RESULT 10
BQ011927
LOCUS      562 bp  mRNA  linear  EST 26-MAR-2002
DEFINITION  UI-1-BC1p-ati-b-01-0-UI.s1 NCI_CGAP_P13 Homo sapiens cDNA clone
            UI-1-BC1p-ati-b-01-0-UI 3', mRNA sequence.
ACCESSION  BQ011927
VERSION    BQ011927.1 GI:19736828
KEYWORDS   EST.
-SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 562)
REFERENCE  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE     Tumor Gene Index
```

JOURNAL
COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Steven Brown

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu

Seq primer: M13 FORWARD

POLYA-yes.

Location/Qualifiers

1. 562

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-1-BC1p-ati-b-01-0-UI"

/clone_lib="NCI_CGAP_P13"

/tissue_type="Placenta"

/dev_stage="8-9 weeks"

/lab_host="DH10B (Life Technologies)"

/note="Organ: Placenta; Vector: pT7T3-Pac (Pharmacia) with
a modified polylinker; Site: 1: EcoR I; Site 2: Not I;
NCI_CGAP_P13 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are GA, AGGA. For
additional information, contact: Bento Soares,
bento-soares@uiowa.edu

TAG_LIB=UI-1-BC1p

TAG_TISSUE=placenta human 8 week

TAG_SEQ=GA

BASE COUNT 118 a 151 c 147 g 146 t

ORIGIN

```
Query Match      42.68; Score 426.4; DB 14; Length 562;
Best Local Similarity 97.5%; Pred. No. 5.8e-88;
Matches 433; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY  1 GTCCTTGGGTAGCATGTTCCATCTCCCTTTATATATGGGGTAAATAGGATAC 60
DB  119 GTCCTTGGGTAGCATGTTCCATCTCTCTTTAGAGAGTGGGGTAAATAGGATAC 178
QY  61 CCCCTCCTCCAGGGGTATCCCTCTTTCTAGGAGACTACCCAAAGCTAGGCCCTTTCTTCCA 130
DB  179 CCCCTCCTCCAGGGGTATCCCTCTTTCTAGGAGACTACCCAAAGCTAGGCCCTTTCTTCCA 238
QY  121 GTGAACGTGCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGCACCACTCT 180
DB  239 GTGAACGTGCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGCACCACTCT 298
QY  181 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGGCGGCTACTGTGTTCCCACTAG 240
DB  299 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGGCGGCTACTGTGTTCCCACTAG 358
QY  241 GAGACACCTTGGCGGGGGCTTTGCTCGCGGGAAGCACGACGAGCGTGGGAGGAGGCC 300
DB  359 GAGACACCTTGGCGGGGGCTTTGCTCGCGGGAAGCACGACGAGCGTGGGAGGAGGCC 418
QY  301 CCTCTGCTGTGTTGTGCCAACAGACACCGCGCTGCCGCTCGGGGTTCCGGCGGCCGGA 360
DB  419 CCTCTGCTGTGTTGTGCCAACAGACACCGCGCTGCCGCTCGGGGTTCCGGCGGCCGGA 478
QY  361 GTCACACATGATGTACACACATGACACAAAGCCGGTGTCTCATTCGACACAGCGTCCG 420
DB  479 GTCACACATGATGTACACACATGACACAAAGCCGGTGTCTCATTCGACACAGCGTCCG 538
QY  421 AGCTGCACAATGTACACACCGGGGT 444
```



```
Seq primer: M13 Forward
POLYA-Yes.
FEATURES
    source
        Location/Qualifiers
            1..556
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="UI-E-EJ0-ai-1-18-0-UI"
                /clone_lib="UI-E-EJ0"
                /tissue_type="fetal eyes, lens, eye anterior segment,
                optic nerve, retina, Retina Foveal and Macular, RPE and
                Choroid"
                /dev_stages="fetal and adult"
                /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                modified polylinker; Site_1: EcoR I; Site_2: Not I;
                UI-E-EJ0 is a subcloned cDNA library constructed
                according to Bonaldo, Lennon and Soares, Genome Research,
                6:791-806, 1996. First strand cDNA synthesis was primed
                with an oligo-dT primer containing a Not I site. Double
                stranded cDNA was ligated to an EcoR I adaptor, digested
                with Not I, and cloned directionally into pT7T3-Pac
                vector. The oligonucleotide used to prime the synthesis of
                first-strand cDNA contains a library tag sequence that is
                located between the Not I site and the (dr)18 tail. The
                sequence tags for this library are: fetal eyes, AGAATCAAGA
                ; lens, CGATTAGCGA; eye anterior segment, AATGCGCAT;
                optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and
                Macular, GTCC; RPE and Choroid, ACCTA. This library was
                created for the program, Gene Discovery in the Visual
                System, supported by National Eye Institute (NEI).
                TAG_LIB=UI-E-EJ0
                TAG_TISSUE=Foveal and Macular Retina
                TAG_SEQ=GTCC"
BASE COUNT      117 a   149 c   145 g   145 t
ORIGIN
Query Match      42.3%; Score 422.6; DB 14; Length 556;
Best Local Similarity 97.9%; Pred. No. 4.3e-87;
Matches 428; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Qy 1 GTCCTGGGTAGCATGATTCATCTCCCTCTTTTATATATATGGGGTAATAGATAC 60
Db 120 GTCCTGGGTAGCATGATTCATCTCCCTCTTTTATATATATGGGGTAATAGATAC 179
Qy 61 CCCCTCCCTCCAGGGGTATCCCTCTTTCTAGGGACCTACCCAAAGCTAGGCCCTTCTTCCA 120
Db 180 CCCCTCCCTCCAGGGGTATCCCTCTTTCTAGGGACCTACCCAAAGCTAGGCCCTTCTTCCA 239
Qy 121 GTGAACGTGTCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGCACCACTCT 180
Db 240 GTGAACGTGTCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGCACCACTCT 299
Qy 181 TCCTTTTATCTCTCAGAGCTGGACAGTGCACCGAGGGCGGTACTGTGTTCCCAAGCTAG 240
Db 300 TCCTTTTATCTCTCAGAGCTGGACAGTGCACCGAGGGCGGTACTGTGTTCCCAAGCTAG 359
Qy 241 GAGACACCTTGGGGGGCTTTGCTCGCGGAGCAGCAGCAGCGCTGGGGAGAGGGGCC 300
Db 360 GAGACACCTTGGGGGGCTTTGCTCGCGGAGCAGCAGCAGCGCTGGGGAGAGGGGCC 419
Qy 301 CCTCTGCTGTGTGTGTCGAACAGCAGCCGCGTGCAGCGCTGGGTTCGGCGGCCGGA 479
Db 420 CCTCTGCTGTGTGTGTCGAACAGCAGCCGCGTGCAGCGCTGGGTTCGGCGGCCGGA 479
Qy 361 GTCACACATGATGTACAGACAATGACAAAGCCGGTGTCTCATTCGACACAGGGTCCG 420
Db 480 GTCACACATGATGTACAGACAATGACAAAGCCGGTGTCTCATTCGACACAGGGTCCG 420
Qy 421 AGCTGCACAATGTCA 437
Db 540 AGCTGCACAATGTCA 556
```

```
RESULT 13
LOCUS      BF061934
DEFINITION 7k69c04.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3480582 3',
            mRNA sequence.
ACCESSION  BF061934
VERSION     BF061934.1
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 530)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            seq primer: -40UP from Gibco
            High quality sequence stop: 459.
FEATURES
    source
        Location/Qualifiers
            1..530
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3480582"
                /clone_lib="NCI_CGAP_GC6"
                /tissue_type="pooled germ cell tumors"
                /lab_host="DH10B"
                /note="Vector: pT7T3-D-Pac (Pharmacia) with a modified
                polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                from the normalized library NCI_CGAP_GC4 was prepared, and
                ss circles were made in vitro. Following HAP purification,
                this DNA was used as tracer in a subtractive hybridization
                reaction. The driver was PCR-amplified cDNAs from a pool
                of 5,000 clones made from the same library (cloneIDs
                1257096-1258631, 1469064-1470983, and 1475592-1476743).
                Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      112 a   145 c   142 g   131 t
ORIGIN
Query Match      41.0%; Score 409.6; DB 12; Length 530;
Best Local Similarity 97.9%; Pred. No. 4.2e-84;
Matches 415; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Qy 1 GTCCTGGGTAGCATGATTCATCTCCCTCTTTTATATATATGGGGTAATAGGATAC 60
Db 107 GTCCTGGGTAGCATGATTCATCTCCCTCTTTTATAGAGATGGGGTAATAGGATAC 166
Qy 61 CCCCTCCTCCAGGGGTATCCCTCTTTCTAGGGACCTACCCAAAGCTAGGCCCTTCTTCCA 120
Db 167 CCCCTCCTCCAGGGGTATCCCTCTTTCTAGGGACCTACCCAAAGCTAGGCCCTTCTTCCA 226
Qy 121 GTGAACGTGTCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGCACCACTCT 180
Db 227 GTGAACGTGTCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGCACCACTCT 286
Qy 181 TCCTTTTATCTCTCAGAGCTGCACAGTGCACCGAGGGCGGTACTGTGTTCCCAAGCTAG 240
Db 287 TCCTTTTATCTCTCAGAGCTGCACAGTGCACCGAGGGCGGTACTGTGTTCCCAAGCTAG 346
Qy 241 GAGACACCTTGGGGGGCTTTGCTCGCGGAGCAGCAGCAGCGCTGGGGAGAGGGGCC 300
Db 347 GAGACACCTTGGGGGGCTTTGCTCGCGGAGCAGCAGCAGCGCTGGGGAGAGGGGCC 406
```

```

QY 301 CCTCTGCTGTGTTTGTGCAACAGCACCGCGCTCGCGTTCGGGTTCCGGCGCGGA 360
DB 407 CCTCTGCTGTGTTGTCACACAGCACCGCGCTCGCGTTCGGGTTCCGGCGCGGA 466
QY 361 GTACACATGATGTCACACACATGACACAAAGCGGTGTCTCATTCGACACACGCTCCG 420
DB 467 GTACACATGATGTCACACACATGACACAAAGCGGTGTCTCATTCGACACACGCTCCG 526
QY 421 AGCT 424
DB 527 AGCT 530

RESULT 14
BM849644/c
LOCUS
DEFINITION K-EST0129836 S21SNU520 Homo sapiens cDNA clone S21SNU520-36-A07 5',
mRNA sequence.
ACCESSION BM849644
VERSION BM849644.1 GI:19206043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 446)
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 36 row: A column: 07
High quality sequence stop: 446.
Location/Qualifiers
1..446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S21SNU520-36-A07"
/clone_lib="S21SNU520"
/sex="F"
/tissue_type="Stomach"
/cell_type="Floating aggregates"
/cell_line="SNU-520"
/lab_host="Top10F"
/notes="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
BASE COUNT 73 a 156 c 151 g 65 t 1 others
ORIGIN

Query Match 40.6%; Score 406; DB 14; Length 446;
Best Local Similarity 99.8%; Pred. No. 2.7e-83;

```

```

Matches 406; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 275 CAGCAGAGCGTGGGAGGAGGCCCCCTCTGCTCTGTTTGTGCCACACAGCACCGCGGC 334
DB 446 CAGCAGAGCGTGGGAGGAGGCCCCCTCTGCTCTGTTTGTGCCACACAGCACCGCGGC 387
QY 335 TGGCGCGTTCGGTTCGGCGCGCGAGTTCACACATGATGTCACAGACATGACACAAGCC 394
DB 386 TGGCGCGTTCGGTTCGGCGCGCGAGTTCACACATGATGTCACAGACATGACACAAGCC 327
QY 395 GGTGTCTCATTCGACACAGCGTCCGAGCTGCACAATGTCACACCGCGGTGCCAAACACT 454
DB 326 GGTGTCTCATTCGACACAGCGTCCGAGCTGCACAATGTCACACCGCGGTGCCAAACACT 267
QY 455 TGGCCCGCGGAGCCCGGCGCTAGCCCTCTGCGCGCGCTCTCCGCTCTCCGGGGAGG 514
DB 266 TGGCCCGCGGAGCCCGGCGCTAGCCCTCTGCGCGCGCTCTCCGCTCTCCGGGGAGG 207
QY 515 TGGCCCGGTTCGGCGCGGAGGCGTGGCGGCGAGCGCGCGCGGGTGGGAGCG 574
DB 206 TGGCCCGGTTCGGCGGCGAGGCGTGGCGGCGAGCGCGCGGGCGGTGGCGAGCG 147
QY 575 GGTGATCTACGGGCGAGCGTGGTGGGTCTACTCGGAGTGAGCGCGCGCGCGAGTT 634
DB 146 GGTGATCTACGGGCGAGCGTGGTGGGTCTACTCGGAGTGAGCGCGCGCGCGAGTT 87
QY 635 CAGCGAGGTTTCAGCCGCATTGTCATTAGGCAAAATGAGCGCGCGCGCTG 681
DB 86 CAGCGAGGTTTCAGCCGCATTGTCATTAGGCAAAATGAGCGCGCGCGCTG 40

RESULT 15
AA404273
LOCUS
DEFINITION zv63008.s1 Soares total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
IMAGE:758319 3', mRNA sequence.
ACCESSION AA404273
VERSION AA404273.1 GI:2058997
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 543)
Hillier, L., Allien, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marta, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie
, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
Contact: Wilson RK
Unpublished (1997)
TITLE
JOURNAL
COMMENT
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 469.
Location/Qualifiers
1..543
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:758319"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/notes="Vector: pTZ3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGGAGCGGCTTAATTTTTTTTTTTTTTTT 3'].

```

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT	117 a	148 c	143 g	134 t	1 others
ORIGIN					
Query Match	39.4%	Score	393.8	DB 9	Length 543
Best Local Similarity	96.6%	Pred. No.	1.1e-80		
Matches 423	Conservative 0	Mismatches 13	Indels 2	Gaps 2	
Qy	1	GTCCCTGGGTAGCATGATACATTTCCATCTTCCTTTTATATATGGGGTAAATAGGATAC	60		
Db	107	GTCCCTGGGTAGCATGATACATTTCCATCTTCCTTTTATAGAGAGTGGGGTAAATAGGATAC	166		
Qy	61	CCCCTCCTCCAGGGTATCCCTCTTCTTAGGGACCTACCCAAAGCTAGGCCCTTCTTCCA	120		
Db	167	CCCCTCCTCCAGGGTATCCCTCTTCTTAGGGACCTACCCAAAGCTAGGCCCTTCTTCCA	226		
Qy	121	GTGAACGTGTCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGSCACAGCTCT	180		
Db	227	GTGAACGTGTCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGSCACAGCTCT	286		
Qy	181	TCCTTTTATCTCTCCAGAGCTGGACAGTGCACACAGGGCGGTACTGTGTTCCCCAGCTAG	240		
Db	287	TCCTTTTATCTCTCCAGAGCTGGACAGTGCACACAGGGCGGTACTGTGTTCCCCAGCTAG	346		
Qy	241	GAGACACCTTGGCGGGGCTTTGCTGCGCGGAAGCACGACAGCGTGGGGA-GGAGGGCC	299		
Db	347	GAGACACCTTGGCGGGGCTTTGCTGCGCGGAAGCACGACAGCGTGGGGA-GGAGGGCC	406		
Qy	300	CCCTCTGCCTGTGTTGTGCCAACAGCACCGCGGTGCGCGGTCCGGCGGCGCGG	359		
Db	407	CCCTCTGCCTGTGTTGTGCCAACAGCACCGCGGTGCGCGGTCCGGCGGCGCGG	465		
Qy	360	AGTCACACATGATGCACAGACAATGACACAGCCGGTGTCTCATTTCCGACACAGCGTCC	419		
Db	466	AGTCACACATGATGCACAGACAATGACACAGCCGGTGTCTCATTTCCGACACAGCGTCC	525		
Qy	420	GAGCTGCACAATGTCACA	437		
Db	526	GAGCTGCACAATGTCACA	543		

Search completed: March 30, 2003, 13:48:06
Job time : 1296.28 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:00:38 ; Search time 2122.53 seconds
(without alignments)
13725.086 Million cell updates/sec

Title: US-09-691-220-3_COPY_5000_6000
Perfect score: 1001
Sequence: 1 tgatctccctccccccacc.....tgggtcacccctcctggctg 1001

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl :

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.ov.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.in.*
- 18: em.mu.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1001	100.0	16913	9	HSA297538	AJ297538 Homo sapi
2	1001	100.0	138999	9	AC090426	AC090426 Homo sapi
3	1001	100.0	158766	2	AC015851	AC015851 Homo sapi
4	1001	100.0	156368	2	AC018629	AC018629 Homo sapi
5	1001	100.0	173441	2	AC080112	AC080112 Homo sapi
6	1001	100.0	188574	2	AC131063	AC131063 Homo sapi
7	1001	100.0	190309	2	AC126392	AC126392 Homo sapi
8	584.2	58.4	13508	6	AX344790	AX344790 Sequence
9	507.2	50.7	13508	6	AX344791	AX344791 Sequence
10	384.2	38.4	516	11	HUMUT5218	L17737 Human STS U
11	146.4	14.6	162393	2	AC111746	AC111746 Rattus no
12	140.4	14.0	231248	2	AL591067	AL591067 Mus muscu
13	60	6.0	914	9	S57794	S57794 RARA=retino
14	59.2	5.9	125020	9	AF429315	AF429315 Homo sapi
15	55	5.5	194929	2	AC119084	AC119084 Rattus no
16	52.2	5.2	173898	2	AC109005	AC109005 Rattus no
17	52	5.2	7218	6	I66494	I66494 Sequence 14
18	51.6	5.2	164520	2	AC020738	AC020738 Homo sapi
19	51	5.1	174031	2	AC118303	AC118303 Rattus no
20	50	5.0	61633	2	AC084075	AC084075 Homo sapi
21	49	4.9	121473	2	AC099003	AC099003 Rattus no
22	48.6	4.9	58928	2	AC096234	AC096234 Rattus no
23	48.4	4.8	26822	2	AC011674	AC011674 Homo sapi
24	47.2	4.7	64055	2	AC111607	AC111607 Rattus no
25	46.8	4.7	171204	2	AC130936	AC130936 Rattus no
26	46.6	4.7	85268	2	AC022648	AC022648 Homo sapi
27	46.6	4.7	117082	2	AC126076	AC126076 Rattus no
28	46.6	4.7	165988	2	AC121212	AC121212 Rattus no
29	46.6	4.7	191277	2	AC128784	AC128784 Rattus no
30	46.4	4.6	130899	2	AC126730	AC126730 Rattus no
31	46.4	4.6	236562	2	AL772338	AL772338 Mus muscu
32	46.2	4.6	168280	2	AC123357	AC123357 Rattus no
33	46	4.6	101509	2	AC027353	AC027353 Homo sapi
34	46	4.6	125020	9	AF429315	AF429315 Homo sapi
35	45.8	4.6	138467	2	AC111697	AC111697 Rattus no
36	45.8	4.6	180758	2	AC128063	AC128063 Rattus no
37	45.8	4.6	196320	9	AL731559	AL731559 Human DNA
38	45.6	4.6	174845	2	AC128072	AC128072 Rattus no
39	45.4	4.5	188133	2	AC112428	AC112428 Rattus no
40	45.2	4.5	68330	2	AC116109	AC116109 Mus muscu
41	45.2	4.5	194776	2	AC129706	AC129706 Rattus no
42	45	4.5	174248	2	AC113654	AC113654 Rattus no
43	45	4.5	183334	2	AC106174	AC106174 Rattus no
44	44.8	4.5	54677	2	AC108286	AC108286 Rattus no
45	44.8	4.5	55365	2	AC095743	AC095743 Rattus no

ALIGNMENTS

RESULT 1
HSA297538
LOCUS HSA297538 Homo sapiens partial RARA gene, intron 2. linear PRI 06-JAN-2001
DEFINITION
ACCESSION AJ297538
VERSION AJ297538.1 GI:12054223
KEYWORDS Rara gene; retinoic acid receptor alpha.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 16913)
AUTHORS Cross,N.C.P. and Reiter,A.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 16913)

AUTHORS Cross, N.C.P.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-2000) Cross N.C.P., Department of Haematology, Imperial College School of Medicine, Hammersmith Hospital, London,

FEATURES Location/Qualifiers

```

source
1. .16913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"

```

gene
1. 16913
17912

intron
1. 16913
/yelles RNA

```
/gene="KARA"  
/number=2
```

```
repeat_region 1638. .1947
/rnt family=
```

```
repeat_region 1949. .2235
              /root family=
```

```
repeat_region 2236. .2536
```

repeat_region	4008.	.4315
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1
20	1	1
21	1	1
22	1	1
23	1	1
24	1	1
25	1	1
26	1	1
27	1	1
28	1	1
29	1	1
30	1	1
31	1	1
32	1	1
33	1	1
34	1	1
35	1	1
36	1	1
37	1	1
38	1	1
39	1	1
40	1	1
41	1	1
42	1	1
43	1	1
44	1	1
45	1	1
46	1	1
47	1	1
48	1	1
49	1	1
50	1	1
51	1	1
52	1	1
53	1	1
54	1	1
55	1	1
56	1	1
57	1	1
58	1	1
59	1	1
60	1	1
61	1	1
62	1	1
63	1	1
64	1	1
65	1	1
66	1	1
67	1	1
68	1	1
69	1	1
70	1	1
71	1	1
72	1	1
73	1	1
74	1	1
75	1	1
76	1	1
77	1	1
78	1	1
79	1	1
80	1	1
81	1	1
82	1	1
83	1	1
84	1	1
85	1	1
86	1	1
87	1	1
88	1	1
89	1	1
90	1	1
91	1	1
92	1	1
93	1	1
94	1	1
95	1	1
96	1	1
97	1	1
98	1	1
99	1	1
100	1	1

repeat_region	4327.4619
---------------	-----------

repeat_region	4669.4975
---------------	-----------

```
repeat_region 4982. 5009
```

```
repeat_region 5011.5287
```

```
repeat_region      5287: .5597  
/ipt_family=
```

```
repeat region
/rpt_family=
5666. .5776
```

```
repeat_region
/rpt_family=
7090. 7148
```

```
repeat region
/rpt_family=
7159. 7326
```

```
repeat region      /rpt_family=
7331 7641
```

```
repeat region      8518      8650
/rpt_family=
```

```

repeat region
0052- 0104
/rpt_family=

```

```

/rpt_family=
10145 1010

```

/note=" (CGGG

2)

Every Match 100.08;

Archives 1001; Conservative

1 TGATCTCCCTCCCCCACCAGT

15105 TGATCTCCCTCCCCCACCAGT

61 AAAGTTAGCACAAAGGGCTG

15165 AAAGTTAGCACAAACAAGGGCTG

127 AGGCCCCCCAACCCCGCCCCC

5225 AGGCCCCCAGCCCCCCCCC
|||||

187

[illegible]

3

THE UNIVERSITY OF CHICAGO

Db	15345	GTGAAGAGCAGACGGCGGTGGGGAGGTCAGGAGAGAATCTGCTGGGCTGGGGATGGTG	15404
Qy	301	GGCATCAACTGTCCTATTTGCTGCAGGCTGGTCTTGGGGCAGGAAGGAGTGGGGGGCCA	360
Db	15405	GGCATCAACTGTCCTATTTGCTGCAGGCTGGTCTTGGGGCAGGAAGGAGTGGGGGGCCA	15464
Qy	361	TAGCAGTGTGGTCAGCCAGGCTGGCCCTGGGAAGTGGTGCCAGGCACCTACTAAGAGCCA	420
Db	15465	TAGCAGTGTGGTCAGCCAGGCTGGCCCTGGGAAGTGGTGCCAGGCACCTACTAAGAGCCA	15524
Qy	421	GGAAAGCCCTGCCAAGGTTGTTGGGCTAGTTCCTCTGTCATCAGCGCCTACGAGCCCCCA	480
Db	15525	GGAAAGCCCTGCCAAGGTTGTTGGGCTAGTTCCTCTGTCATCAGCGCCTACGAGCCCCCA	15584
Qy	481	CTGTCTCTGCAGTAAGGGGGAGGGTGTAGCACATAGTCAGCCCCCTGGTGTCCCATG	540
Db	15585	CTGTCTCTGCAGTAAGGGGGAGGGTGTAGCACATAGTCAGCCCCCTGGTGTCCCATG	15644
Qy	541	CTTCCCTCTCTGTGCCCAATTTTAGGGCCATGTGATTTGGGGCTATGTGACTCATGTC	600
Db	15645	CTTCCCTCTCTGTGCCCAATTTTAGGGCCATGTGATTTGGGGCTATGTGACTCATGTC	15704
Qy	601	TGTAAGTGCTTTGGGCCAGGAGCTGTGGCACCTTTTAAATGCCAGCAGTCTCATGTGCC	660
Db	15705	TGTAAGTGCTTTGGGCCAGGAGCTGTGGCACCTTTTAAATGCCAGCAGTCTCATGTGCC	15764
Qy	661	GGAGTTTGGGGTAGGGCTAGGTAGTGTGGAATATGGGAGGAGCAGGATCTGTCTA	720
Db	15765	GGAGTTTGGGGTAGGGCTAGGTAGTGTGGAATATGGGAGGAGCAGGATCTGTCTA	15824
Qy	721	CCTAGGGAGGCATCCTCATCCATCCTTGGCCCTTGGACAAGAGACTTCAACCTTGGTAGG	780
Db	15825	CCTAGGGAGGCATCCTCATCCATCCTTGGCCCTTGGACAAGAGACTTCAACCTTGGTAGG	15884
Qy	781	GGCCTCAGGACATGCTCGTGGCCCCCTTGGGAATCTGGGATGTCTGGTCAATAGTTCT	840
Db	15885	GGCCTCAGGACATGCTCGTGGCCCCCTTGGGAATCTGGGATGTCTGGTCAATAGTTCT	15944
Qy	841	TATCTTGACCCACACCCCTTAGCTGCCAGGCTTTGGACATGGATAGCCCTACCCAAC	900
Db	15945	TATCTTGACCCACACCCCTTAGCTGCCAGGCTTTGGACATGGATAGCCCTACCCAAC	16004
Qy	901	CCAGCCCTGTTCTGCCTACAGTGTAGTGGCATGGAGCCACAGACACTGGGAGGATTTGGCCA	960
Db	16005	CCAGCCCTGTTCTGCCTACAGTGTAGTGGCATGGAGCCACAGACACTGGGAGGATTTGGCCA	16064
Qy	961	GTGAGGGCTGCCCTGCTGTCTGGGTACACCCCTCTGGGCTG	1001
Db	16065	GTGAGGGCTGCCCTGCTGTCTGGGTACACCCCTCTGGGCTG	16105

RESULT 2	
AC090426/c	
LOCUS	AC090426
DEFINITION	Homo sapiens chromosome 17 clone 205m17 map 17q21.1, complete sequence.
ACCESSION	AC090426
VERSION	AC090426.1
KEYWORDS	HTG.
SOURCE	Homo sapiens
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 138999)
AUTHORS	Gu,B., Xiong,H., Zhou,Y., Chen,B., Lu,L., Zhong,M., Yin,H., Huang,W., Ren,S., Chen,S.F., Chen,Z. and Fu,G.
TITLE	Variant-type PML-RAR(alpha) fusion transcript in acute promyelocytic leukemia: use of a cryptic coding sequence from intron 2 of the RAR(alpha) gene and identification of a new clinical subtype resistant to retinoic acid therapy
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (11), 7640-7645 (2002)
MEDLINE	22028997
PUBMED	12032336

```

REFERENCE      2 (bases 1 to 138999)
AUTHORS        Gu.B., Xiong.H., Zhou.Y., Chen.B., Lu.L., Zhong.M., Yin.H.,
                Huang.W., Ren.S., Chen.S.F., Chen.Z. and Fu.G.
TITLE          Direct Submission
JOURNAL        Shanghai, Shanghai, Shanghai 201203, P.R.China
COMMENT        -----Genome Center-----
                Center:Chinese National Human Genome Center at Shanghai
                Center code:CHGC Website: http://www.chgc.sh.cn Contact:
                fugang@chgc.sh.cn.
FEATURES
source
1. .138999
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="17"
   /map="17q21.1"
   /clone="205ml7"
   /map="387. .483
repeat_region /rpt_family="SINE/MIR"
               /rpt_unit=MIR
repeat_region 725. .1032
               /rpt_family="SINE/Alu"
               /rpt_unit=AluB
repeat_region 1033. .1198
               /rpt_family="SINE/Alu"
repeat_region 1217. .1514
               /rpt_family="SINE/Alu"
               /rpt_unit=AluSp
repeat_region complement(1534. .1754)
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
repeat_region 2291. .2366
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
repeat_region complement(2607. .2665)
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
repeat_region 3651. .3905
               /rpt_family="LINE/L1"
               /rpt_unit=L1ME2
repeat_region 3977. .4029
               /rpt_family="LINE/L1"
               /rpt_unit=L1ME2
repeat_region 4030. .4213
               /rpt_family="LTR/MaLR"
               /rpt_unit=MSTD
repeat_region 4214. .4509
               /rpt_family="SINE/Alu"
               /rpt_unit=AluB
repeat_region 4514. .4811
               /rpt_family="SINE/Alu"
               /rpt_unit=AluSx
repeat_region 4812. .5018
               /rpt_family="LTR/MaLR"
               /rpt_unit=MSTD
repeat_region 5019. .5102
               /rpt_family="LINE/L1"
               /rpt_unit=L1ME2
repeat_region 5236. .5443
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
repeat_region 5467. .5699
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
repeat_region 6547. .6708
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
repeat_region 6827. .6917
               /rpt_family="Low_complexity"
               /rpt_unit=GA-rich
repeat_region 7733. .7779
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
8268. .8411
/rpt_family="Simple_repeat"
/rpt_unit=(TGGG)n
9224. .9532
/rpt_family="SINE/Alu"
/rpt_unit=AluY
9546. .9639
/rpt_family="Low_complexity"
/rpt_unit=CT-rich
10025. .10426
/rpt_family="LTR/Retroviral"
/rpt_unit=MLT2A
10433. .10471
/rpt_family="Simple_repeat"
/rpt_unit=(TA)n
10472. .10527
/rpt_family="LTR/Retroviral"
/rpt_unit=MLT2A
complement(10675. .10962)
/rpt_family="SINE/Alu"
/rpt_unit=AluY
11013. .11321
/rpt_family="SINE/Alu"
/rpt_unit=AluY
complement(11840. .12192)
/rpt_family="LTR/MaLR"
12729. .12835
/rpt_family="SINE/Alu"
/rpt_unit=FLAM_A
complement(12841. .12891)
/rpt_family="LINE/L2"
/rpt_unit=L2
complement(14228. .14548)
/rpt_family="SINE/Alu"
/rpt_unit=AluY
complement(14618. .14701)
/rpt_family="SINE/MIR"
14953. .15255
/rpt_family="SINE/Alu"
/rpt_unit=AluJo
15311. .15473
/rpt_family="Simple_repeat"
/rpt_unit=(TA)n
15669. .15981
/rpt_family="SINE/Alu"
/rpt_unit=AluSx
16316. .16353
/rpt_family="Simple_repeat"
/rpt_unit=(TAAA)n
complement(16863. .16910)
/rpt_family="SINE/MIR"
17026. .17324
/rpt_family="SINE/Alu"
/rpt_unit=AluSg
17337. .17618
/rpt_family="SINE/Alu"
/rpt_unit=AluSx
17619. .18024
/rpt_family="LTR"
/rpt_unit=LTR54
complement(18026. .18086)
/rpt_family="SINE/MIR"
/rpt_unit=MIR
complement(18292. .18607)
/rpt_family="SINE/Alu"
/rpt_unit=AluYb8
complement(18948. .19264)
/rpt_family="SINE/Alu"
/rpt_unit=AluSx
complement(20012. .20168)

```



```
QY 901 CCAGCCCTGTTCTGCTACAGTGTGGGCGATGGGACAGACACTGGGAGGATTGGGCA 960
|||||
Db 78508 CCAGCCCTGTTCTGCTACAGTGTGGGCGATGGGACAGACACTGGGAGGATTGGGCA 78567
|||||
QY 961 GTGAGGCGTCCCTGCTGCTGTGCTGGGTTCACCCCTCTGCTGCTG 1001
|||||
Db 78568 GTGAGGCGTCCCTGCTGCTGTGCTGGGTTCACCCCTCTGCTGCTG 78608
|||||

RESULT 4
AC018629 166368 bp DNA linear HTG 19-AUG-2002
LOCUS Homo sapiens chromosome 17 clone RP11-5809 map 17, *** SEQUENCING
DEFINITION IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC018629
VERSION AC018629.10 GI:22297481
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166368)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-5809
Unpublished
2 (bases 1 to 166368)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Farrell,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lander,E., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H.,
O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A.,
Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Vassiliev,H., Viel,R., Vo.A., Wu.X., Wyman,D., Ye.W.J., Zimmer,A.
and Zody,M.
Direct Submission
Submitted (14-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 166368)
Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgaiter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo.A., Wilson,B., Wu.X., Wyman,D., Young,G., Zainoun,J.,
Zenbek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 19, 2002 this sequence version replaced gi:22123724.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L924
Center clone name: 58_Q_9
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 83845: contig of 83845 bp in length
* 83846 83945: gap of 100 bp
* 83946 166368: contig of 82423 bp in length.
FEATURES
Location/Qualifiers
1..166368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-5809"
/clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 39988 a 40877 c 42066 g 43337 t 100 others
ORIGIN
Query Match 100.0% Score 1001; DB 2; Length 166368;
Best Local Similarity 100.0%; Pred. No. 3.3e-267;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGATCTCCCTCCCTCCCTCCCTCCCTCCCTCTGATGTTGTTACTGCTTTTACGTTTGGAA 60
Db 1549 TGATCTCCCTCCCTCCCTCCCTCCCTCCCTCTGATGTTGTTACTGCTTTTACGTTTGGAA 1608
QY 61 AAAGTTAGCACAAACAAAGGCTGTTTGGCTCACCCCTCTGCTCCCTCCCTCCCTCACCC 120
Db 1609 AAAGTTAGCACAAACAAAGGCTGTTTGGCTCACCCCTCTGCTCCCTCCCTCACCC 1668
QY 121 AGGCCCCCAACCCCGCCCCCCCCAGCAGCTGTTCTCAGGCGCTCTCAGCGTGTCTGATTTG 180
Db 1669 AGGCCCCCAACCCCGCCCCCCCCAGCAGCTGTTCTCAGGCGCTCTCAGCGTGTCTGATTTG 1728
QY 181 CTTGCTGCTGGCTGGGAGGATGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
Db 1729 CTTGCTGCTGGCTGGGAGGATGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1788
QY 241 GTGAAGAGCAGACGCGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 1789 GTGAAGAGCAGACGCGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1848
QY 301 GGCATCAACTGTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Db 1849 GGCATCAACTGTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1908
QY 361 TAGCAGTGTGTCAGCCAGGCTGGCTGGGAGTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGG 420
Db 1909 TAGCAGTGTGTCAGCCAGGCTGGCTGGGAGTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGG 1968
QY 421 GGAAGCCCTGCAAGGTTGTTGGCTAGTCCCTCATCAGCCCGCTAGCAGCCCGCTAGCAGCCCG 480
Db 1969 GGAAGCCCTGCAAGGTTGTTGGCTAGTCCCTCATCAGCCCGCTAGCAGCCCGCTAGCAGCCCG 2028
QY 481 CTGCTGCTCAGGTAAAGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
Db 2029 CTGCTGCTCAGGTAAAGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2088
QY 541 CTTCTCTCTCTGTCGCCCAATTTTAGGGCCCATGTGATTTGGGGCTATGTGACTCATGTC 600
Db 2089 CTTCTCTCTCTGTCGCCCAATTTTAGGGCCCATGTGATTTGGGGCTATGTGACTCATGTC 2148
```

QY 601 TGTAAAGGTCTTGGGCGAGAGCTGTGGCACCCTTTAAATGCGAGCCAGTCTCATGTGCC 660
|||||
Db 2149 TGTAAAGGTCTTGGGCGAGAGCTGTGGCACCCTTTAAATGCGAGCCAGTCTCATGTGCC 2208
QY 661 GGAGTTTGGGTPAGGCTAGGTAGGATGTGGAATATGGAGAGAGCAGGATCTCTCTA 720
|||||
Db 2209 GGAGTTTGGGTPAGGCTAGGTAGGATGTGGAATATGGAGAGAGCAGGATCTCTCTA 2268
QY 721 CCTAGGGAGGATCTCATCATCTCTTGGCCCTTGGACAGAGAACTTGACGTTGGTAGG 780
|||||
Db 2269 CCTAGGGAGGATCTCATCATCTCTTGGCCCTTGGACAGAGAACTTGACGTTGGTAGG 2328
QY 781 GSCCTCAGGACCATGCTGGTGCGCCCTTGGGAATCTGGGATGTCTCTGTCATCTTCT 840
|||||
Db 2329 GSCCTCAGGACCATGCTGGTGCGCCCTTGGGAATCTGGGATGTCTCTGTCATCTTCT 2388
QY 841 TATCTTGGACCAACACCCCTTACGCTGCCAGGCTTGGACATGGATAGCCCTTACCCAAC 900
|||||
Db 2389 TATCTTGGACCAACACCCCTTACGCTGCCAGGCTTGGACATGGATAGCCCTTACCCAAC 2448
QY 901 CGAGCCCTGTTCTGCTACAGTGTGGCATGGAGCCAGACACTGGGGAGGATTTGGCCA 960
|||||
Db 2449 CGAGCCCTGTTCTGCTACAGTGTGGCATGGAGCCAGACACTGGGGAGGATTTGGCCA 2508
QY 961 GTGAGGGTGGCCCTGCTGCTGGGTGTCACCCCTCTGGGTG 1001
|||||
Db 2509 GTGAGGGTGGCCCTGCTGCTGGGTGTCACCCCTCTGGGTG 2549

RESULT 5
AC080112/c
LOCUS AC080112 173441 bp DNA linear HTG 14-AUG-2002
DEFINITION Homo sapiens chromosome 17 clone CTD-2267D19 map 17, *** SEQUENCING
IN PROGRESS ***, 3 unordered pieces.
AC080112
VERSION AC080112.6 GI:22218574
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choepe, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
FitzHugh, W., Gage, D., Galagan, J., Gardyna, J., Ginde, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Larkoc, K.,
Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougnaz, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-SEP-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 173441)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,

Camarata, J., Chang, J., Chazaro, B., Choepe, Y., Collymore, A.,
Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (14-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 14, 2002 this sequence version replaced gi:22123315.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11030
Center clone name: 2267_D_19

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2128: contig of 2128 bp in length
* 2129 2228: gap of 100 bp
* 2229 157931: contig of 155703 bp in length
* 157932 158031: gap of 100 bp
* 158032 173441: contig of 15410 bp in length.

Location/Qualifiers
1. 173441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="CTD-2267D19"
/clone_lib="CITD1 Human BAC"
BASE COUNT 46772 a 43344 c 40918 g 42114 t 293 others
ORIGIN

Query Match 100.0%; Score 1001; DB 2; Length 173441;
Best Local Similarity 100.0%; Pred. No. 3.3e-267;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGATCTCCCTCCCGCCAGCTGTGGATGTCTATTGTACTGCTTTTACGCTTGTGAA 60
|||||
Db 77238 TGATCTCCCTCCCGCCAGCTGTGGATGTCTATTGTACTGCTTTTACGCTTGTGAA 77179
QY 61 AAAGTTAGCACAAACAAGGCTGCTTTGTGCTCACCCCTCTGCTCCCTGCCTCACCC 120
|||||
Db 77178 AAAGTTAGCACAAACAAGGCTGCTTTGTGCTCACCCCTCTGCTCCCTGCCTCACCC 77119
QY 121 AGGCCCCCCAACCCCGCCCCCAGCAGCTGTTCTCAGGCTCTCAGCCTCTCGATTG 180
|||||
Db 77118 AGGCCCCCCAACCCCGCCCCCAGCAGCTGTTCTCAGGCTCTCAGCCTCTCGATTG 77059
QY 181 CTTGTCTGGCCTGGGGAGAAATGAGTGGGAGAAACACAGCCGAGGTGGTGGTGGGA 240

Db	77058	CTTGCTGGCGCTGGGAGAAATGAGTGGGAGAAACACAGCCAGCGGATGGTGTGGGA	76999
Qy	241	GTGAAGACGACAGCGCGTGGGAGGTGACGAGAGAATCTGCTGGCTGGGATGGGTG	300
Db	76998	GTGAAGACGACAGCGCGTGGGAGGTGACGAGAGAATCTGCTGGCTGGGATGGGTG	76939
Qy	301	GGCATCAACTGTCCCAATGTCTGACAGCGTGTGGGAGAGGATGGGGGCCA	360
Db	76938	GGCATCAACTGTCCCAATGTCTGACAGCGTGTGGGAGAGGATGGGGGCCA	76879
Qy	361	TAGCAGTGTGCTGACAGCCAGCGCTGGCCCTGGGAAGTGTGCCCAGGCACCTACTAAGACCA	420
Db	76878	TAGCAGTGTGCTGACAGCCAGCGCTGGCCCTGGGAAGTGTGCCCAGGCACCTACTAAGACCA	76819
Qy	421	GGAAAGCCCTGCCAAGGTGTGTGGCCCTAGTTCCTGTCTCATCAGCCGCTAGCAGCCGCCA	480
Db	76818	GGAAAGCCCTGCCAAGGTGTGTGGCCCTAGTTCCTGTCTCATCAGCCGCTAGCAGCCGCCA	76759
Qy	481	CTGTGTCTGCAGGTAAAGGGGGAGGGTGTAGCACATAGTACGCCCTGTGTTCCTCATG	540
Db	76758	CTGTGTCTGCAGGTAAAGGGGGAGGGTGTAGCACATAGTACGCCCTGTGTTCCTCATG	76699
Qy	541	CTTCTCTTCTCTGTGCCCCCAATTTTAGGCCATCTGATTTGGGGCTATGTGACATCATGTC	600
Db	76698	CTTCTCTTCTCTGTGCCCCCAATTTTAGGCCATCTGATTTGGGGCTATGTGACATCATGTC	76639
Qy	601	TGTAAGTGTCTTGGGCCAGGAGCTGTGGGCAACCTTTAAATGCCAGCAGTCTCATGTGCC	660
Db	76638	TGTAAGTGTCTTGGGCCAGGAGCTGTGGGCAACCTTTAAATGCCAGCAGTCTCATGTGCC	76579
Qy	661	GGAGTTTGGGTGAGGCTAGTGTAGGATTTGTGGAATATGGGAGGAGCAGGATCTGTCTA	720
Db	76578	GGAGTTTGGGTGAGGCTAGTGTAGGATTTGTGGAATATGGGAGGAGCAGGATCTGTCTA	76519
Qy	721	CCTAGGAGGACATCCTCATCCATCCTTGGCCCTGGCAAGAGAACTTGAACGTTGGTAGG	780
Db	76518	CCTAGGAGGACATCCTCATCCATCCTTGGCCCTGGCAAGAGAACTTGAACGTTGGTAGG	76459
Qy	781	GGCCTCAGGAGCATGCTGCGTGGCCCTTGGGAATCTGGGAATCTGCTGGTCATAGTTCT	840
Db	76458	GGCCTCAGGAGCATGCTGCGTGGCCCTTGGGAATCTGGGAATCTGCTGGTCATAGTTCT	76399
Qy	841	TATCTTGACCCAAACACCCCTTAGCTGCCAGGCTTTGGACATGGATAGCCCTACCCAAAC	900
Db	76398	TATCTTGACCCAAACACCCCTTAGCTGCCAGGCTTTGGACATGGATAGCCCTACCCAAAC	76339
Qy	901	CCAGCCCTGTTCTGCCCTACAGTGATGGCATGGGACGACAGACTGGGGAGGATTTGGCCA	960
Db	76338	CCAGCCCTGTTCTGCCCTACAGTGATGGCATGGGACGACAGACTGGGGAGGATTTGGCCA	76279
Qy	961	GTGAGGCTGCCCTGCTGCTGCGGTCAACCCCTCCTGGCTG	1001
Db	76278	GTGAGGCTGCCCTGCTGCTGCGGTCAACCCCTCCTGGCTG	76238
RESULT 6			
AC131063/c			
LOCUS			
DEFINITION	AC131063	188574 bp	DNA linear
SEQUENCE			HTG 16-AUG-2002
ACCESSION			
VERSION	AC131063		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 188574)		
JOURNAL	Birren, B., Nusbaum, C. and Lander, E.		
REFERENCE	2 (bases 1 to 188574)		

[illegible]

ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE	1 (bases 1 to 914)

of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 194929)
 Worley, K.C.
 Direct Submission
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20303396.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GTMP
 Center clone name: CH230-112G1
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 128159 bases at least Q40
 Consensus quality: 135188 bases at least Q30
 Consensus quality: 139965 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 71 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1036: contig of 1036 bp in length
 1037 1136: gap of unknown length
 1137 2399: contig of 1263 bp in length
 2400 2499: gap of unknown length
 2500 3523: contig of 1024 bp in length
 3524 3623: gap of unknown length
 3624 5005: contig of 1382 bp in length
 5006 5105: gap of unknown length
 5106 6355: contig of 1250 bp in length
 6356 7907: contig of 1452 bp in length
 7908 8007: gap of unknown length
 8008 9317: contig of 1310 bp in length
 9318 9417: gap of unknown length
 9418 10644: contig of 1227 bp in length
 10645 10744: gap of unknown length
 10745 11925: contig of 1181 bp in length
 11926 12025: gap of unknown length
 12026 13524: contig of 1499 bp in length
 13525 14677: contig of 1053 bp in length
 14678 14777: gap of unknown length
 14778 15855: contig of 1078 bp in length
 15856 15955: gap of unknown length
 15956 17320: contig of 1365 bp in length
 17321 17421: gap of unknown length
 17422 18455: contig of 1035 bp in length
 18456 18555: gap of unknown length
 18556 20137: contig of 1582 bp in length
 20138 20237: gap of unknown length
 20238 21836: contig of 1599 bp in length
 21837 21936: gap of unknown length
 21937 23431: contig of 1495 bp in length
 23432 23531: gap of unknown length
 23532 24970: contig of 1439 bp in length
 24971 25070: gap of unknown length
 25071 26350 26449: gap of unknown length

26450 27905: contig of 1456 bp in length
 27906 28005: gap of unknown length
 28006 29330: contig of 1325 bp in length
 29331 29430: gap of unknown length
 29431 30576: contig of 1146 bp in length
 30577 30676: gap of unknown length
 30677 32160: contig of 1484 bp in length
 32161 32260: gap of unknown length
 32261 33426: contig of 1166 bp in length
 33427 33526: gap of unknown length
 33527 35660: contig of 2134 bp in length
 35661 35760: gap of unknown length
 35761 37766: contig of 2006 bp in length
 37767 37866: gap of unknown length
 37867 39093: contig of 1227 bp in length
 39094 39193: gap of unknown length
 39194 41094: contig of 1901 bp in length
 41095 41194: gap of unknown length
 41195 42396: contig of 1202 bp in length
 42397 42496: gap of unknown length
 42497 43637: contig of 1141 bp in length
 43638 43737: gap of unknown length
 43738 45693: contig of 1956 bp in length
 45694 45793: gap of unknown length
 45794 47577: contig of 1784 bp in length
 47578 47677: gap of unknown length
 47678 49368: contig of 1691 bp in length
 49369 49468: gap of unknown length
 49469 50928: contig of 1460 bp in length
 50929 51028: gap of unknown length
 51029 53617: contig of 2589 bp in length
 53618 53717: gap of unknown length
 53718 55954: contig of 2237 bp in length
 55955 56054: gap of unknown length
 56055 58165: contig of 2111 bp in length
 58166 58265: gap of unknown length
 58266 60797 60896: contig of 2531 bp in length
 60897 63491: contig of 2595 bp in length
 63492 63591: gap of unknown length
 63592 65691: contig of 2100 bp in length
 65692 65791: gap of unknown length
 65792 67770: contig of 1979 bp in length
 67771 70084: contig of 2214 bp in length
 70085 70184: gap of unknown length
 70185 72274: contig of 2090 bp in length
 72275 72374: gap of unknown length
 72375 74673: contig of 2299 bp in length
 74674 74773: gap of unknown length
 74774 77271: contig of 2498 bp in length
 77272 77371: gap of unknown length
 77372 79490: contig of 2119 bp in length
 79491 79590: gap of unknown length
 79591 81823: contig of 2233 bp in length
 81824 81923: gap of unknown length
 81924 83627: contig of 1704 bp in length
 83628 83727: gap of unknown length
 83728 85740: contig of 2013 bp in length
 85741 88540: gap of unknown length
 88541 88656: contig of 2816 bp in length
 88657 88756: gap of unknown length
 88757 91337: contig of 2581 bp in length
 91338 91437: gap of unknown length
 91438 93524: contig of 2487 bp in length
 93525 94024: gap of unknown length
 94025 96899: gap of unknown length
 96899 99470: contig of 2472 bp in length

Query Match 5.5%; Score 55; DB 2; Length 194929;
 Best Local Similarity 45.8%; Pred. No. 0.00028;
 Matches 162; Conservative 0; Mismatches 191; Indels 1; Gaps 1;

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 07:57:38 : Search time 196.299 seconds
(without alignments)
11483.757 Million cell updates/sec

Title: US-09-691-220-3_COPY_5000_6000
Perfect score: 1001
Sequence: 1 tgatccctcccccacc.....tgggtaccctccctggctg 1001

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues 4370478

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
N_Geneseq_101002.*			
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*		
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*		
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*		
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*		
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*		
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*		
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*		
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*		
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*		
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*		
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*		
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*		
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*		
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*		
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*		
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*		
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*		
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*		
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*		
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*		
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*		
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*		
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*		
24:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	20512	24	Genomic DNA encodi
2	584.2	58.4	13508	24	Human 'chemically m
3	507.2	50.7	13508	24	Human chemically m
4	44.6	4.5	549	13	AAQ20379
5	44.6	4.5	549	13	AAQ20467
6	42.6	4.3	357	22	AAI81967
7	42.6	4.3	34980	22	AAH68527
8	42.2	4.2	30032	22	ABAL7086
9	42	4.2	37664	22	AAK83781

C 10	42	4.2	61710	22	AAK83782
C 11	41.4	4.1	9359	24	ABK10687
C 12	41.4	4.1	9359	24	ABK11039
C 13	39.6	4.0	1894	13	AAQ20466
C 14	39.6	4.0	1894	18	AAAT71220
C 15	39.2	3.9	5786	22	AAAS29976
C 16	38.4	3.8	1407	21	AAZ89275
C 17	38.4	3.8	2237	22	AAAS30201
C 18	38.4	3.8	2378	21	AAAD01016
C 19	38.4	3.8	2620	21	AAAS3824
C 20	38.4	3.8	3879	21	AAZ89276
C 21	38.2	3.8	6865	22	AAAF80287
C 22	38.2	3.8	10003	22	AAAF80295
C 23	38.2	3.8	114793	22	AAAD08215
C 24	38	3.8	1337	20	AAAZ17263
C 25	38	3.8	1501	21	AAAC87194
C 26	38	3.8	1501	21	AAAC89325
C 27	38	3.8	4032	19	AAV23237
C 28	37.6	3.8	410	22	AAI82055
C 29	37.6	3.8	600	24	ABQ52496
C 30	37.6	3.8	600	24	ABQ52497
C 31	37.6	3.8	10732	21	AAAI0594
C 32	37.4	3.7	5121	22	AAAK89337
C 33	37.4	3.7	5864	18	AAAT72812
C 34	37.4	3.7	53522	24	AAAD30228
C 35	37.4	3.7	53526	19	AAAT94101
C 36	37.4	3.7	53577	17	AAAT18551
C 37	37.4	3.7	53577	19	AAAT94108
C 38	37.2	3.7	840	24	ABQ35494
C 39	37.2	3.7	840	24	ABQ35495
C 40	37.2	3.7	993	24	ABK35988
C 41	37.2	3.7	2042	24	ABN95855
C 42	37	3.7	801	22	AAH99234
C 43	37	3.7	1521	22	AAAD08644
C 44	37	3.7	1578	23	AAAS85330
C 45	37	3.7	2380	22	AAAS14817

ALIGNMENTS

RESULT 1	
ID	AAK38339 standard; DNA; 20512 BP.
XX	AAK38339;
AC	
XX	
DT	15-AUG-2002 (first entry)
DE	Genomic DNA encoding the human nuclear hormone receptor protein.
XX	
KW	Human nuclear hormone receptor; therapeutic agent; gene therapy;
KW	Immune response; chromosome 17; single nucleotide polymorphism;
KW	gene; ds.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1684..15208
FT	/tag= a
FT	/product= "Human nuclear hormone receptor protein"
FT	/tag= b
FT	1684..1846
FT	/number= 1
FT	1847..6808
FT	/tag= c
FT	/number= 1
FT	6809..6957
FT	/tag= d
FT	/number= 2
FT	6958..8277
FT	/tag= e
FT	/number= 2

Db 5600 TGTAAGGTGCTTGGGCGAGAGCTGTGGGCAACCTTTAAATGCCAGCAGTCTCATGTGCC 5659
QY 661 GGAGTTTGGGGTAGGCTAGGTAGGATTTGGGAATATGGAGAGGAGGAGGATCTGCTA 720
Db 5660 GGAGTTTGGGGTAGGCTAGGTAGGATTTGGGAATATGGAGAGGAGGAGGATCTGCTA 5719
QY 721 CTTAGGAGGCAATCTATCCATCCCTTGGCCCTGGACAGAGAACTTGAACGTTGGTAGG 780
Db 5720 CTTAGGAGGCAATCTATCCATCCCTTGGCCCTGGACAGAGAACTTGAACGTTGGTAGG 5779
QY 781 GGCTCAGCAGCATGTGCTGGCCCTTGGGAATCTGGGATCTGCTGTCTATAGTTCT 840
Db 5780 GGCTCAGCAGCATGTGCTGGCCCTTGGGAATCTGGGATCTGCTGTCTATAGTTCT 5839
QY 841 TATCTTGCAACCAACCCCTTACTGCCAGGCTTTGGACATGATGATCCCTTACCACAAC 900
Db 5840 TATCTTGCAACCAACCCCTTACTGCCAGGCTTTGGACATGATGATCCCTTACCACAAC 5899
QY 901 CCAGCCCTTCTTGCCTACAGTATGGGATGGGATGGGAGGATTTGGGCA 960
Db 5900 CCAGCCCTTCTTGCCTACAGTATGGGATGGGATGGGAGGATTTGGGCA 5959
QY 961 GTGAGGGCTGCCCTCTCTCTGGGTACACCCCTCTCTGGGTG 1001
Db 5960 GTGAGGGCTGCCCTCTCTCTGGGTACACCCCTCTCTGGGTG 6000

RESULT 2

ABN80198
ID ABN80198 standard; DNA; 13508 BP.
AC ABN80198;
XX
XX
XX 15-JUL-2002 (first entry)
XX Human chemically modified disease associated gene SEQ ID NO 215.
DE Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
XX WO200200927-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07536.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A. Piepenbrock C, Berlin K;
XX WPI; 2002-130908/17.
XX
XX Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a
PT sequence of a segment of chemically pretreated DNA of genes associated
PT with development
XX
XX Claim 1; SEQ ID NO 215; 27pp; English.
XX
XX The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in
CC the specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
CC sequences (ABN7984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with

CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs).
CC Note: The sequence data for this patent did not form part of the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.

XX Sequence 13508 BP; 2350 A; 286 C; 4030 G; 6841 T; 1 other;

Query Match 58.4%; Score 584.2; DB 24; Length 13508;
Best Local Similarity 74.7%; Pred. No. 2.3e-152;
Matches 733; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

QY 21 AGTCTGGATGTCTATTGTTACTGCTTTTACGTTCTTGAAAAAGTTAGCAACAAGGG 80
Db 11719 AGTTGGATGTTTATTGTTATTGTTTACGTTTGGAAAAAGTTAGTAATAAAGGG 11778
QY 81 CTGCTTTGGCTCACCCCTCTGCTGCGCTCACCCAGGCCCCCAACCCGCCCC 140
Db 11779 TTGTTTGGCTTATTATTTTGTGTTTATTTAGTTTAAATTCGTTTT 11838
QY 141 CCCAGCAGCTGTTCTCAGGCTCTCAGCCTGCTGATTTGCTGCGCTGGGAGAA 200
Db 11839 TTTAGTAGTTGTTTATAGGTTTATGTTGTTGTTGTTGTTGGGGAGAA 11898
QY 201 TGAGTGGGAGAAACCCAGGCGAGTGTGTTGGAGTGAAGACAGCAGCGGTG 260
Db 11899 TGAGTGGGAGAAATAGCTTAGGTAGTTGGTGTGAGTGAAGACAGCGCGGTG 11958
QY 261 GGAGGTGAGGAGAGATCTGCTGGGCTGGGATGTTGGGATCACTCTCCATTCG 320
Db 11959 GGGAGGTAGGAGAGAAATTTGTTGGGTGGGATGTTGGGTATTAATTTATTGT 12018
QY 321 TGCAGCTGCTCTGCGGAGGAGGATGGGGGCCATAGCAGTCTGCTGAGCCAG 380
Db 12019 TGTAGGTGTTGTTGGGTAGGAGGATGGGGGTATAGTAGTGTGTTAGTTAG 12078
QY 381 GCTGGCTTGGAAAGTGTGCCAGGCACTACTAAGACCCAGGAAGCCCTCCCAAGTTG 440
Db 12079 GTTGGTTGGGAAGTGTGTTAGTATTAAGAGTTAGGAAGGTTTGTGAAGTTG 12138
QY 441 TTGGCTAGTCTCCCTGTCATCAGCCGCTAGCAGCCCCCTGCTGCTGAGGTAAAGGG 500
Db 12139 TTGGTTAGTTTCTGTTAGTCTGTTAGTATTTATTTGTTGTTGTTAGGTAAAGGG 12198
QY 501 GGAGGTGTTAGCAGATAGTACCCCTGTTGTTCCCATGCTCTCTCTGTCGCCCA 560
Db 12199 GGAGGTGTTAGTATAGTATGTTTGTGTTTATGTTTATGTTTGTGTTTAT 12258
QY 561 ATTTTAGGCGCATGTGATTTGGGCTATGACTCATGCTCTAAGCTGCTTGGCGCAGG 620
Db 12259 ATTTTAGGCTATGCTATGTTGGGCTATGATTTATGTTTGAAGGTGTTGGGTAGG 12318
QY 621 AGCTGGGCACTTTAAATGCCAGCAGTCTCATGCTGCCGGAGTTTGGGTAGGGGTAG 680
Db 12319 AGTTGTTGGTATTTTAAATGTTAGTTAGTTTATGTTGTCGAGTTTGGGTAGGTTAG 12378
QY 681 GTAGGATTTGGAATATGGAGGAGGATGCTGCTACCTAGGAGGATCCTCATC 740
Db 12379 GTAGGATTTGGAATATGGAGGAGGATGCTGCTACCTAGGAGGATCCTCATC 12438
QY 741 CATCTTGGCCCTGGACAGAGAGACTTGAACGTTGTTAGGGGCTCAGGAGGATGCTCG 800
Db 12439 TATTTTGGTGGATTAAGAAATTTGAACGTTGTTAGGGGTTTAAAGCAGATGTTGCG 12498

Db 910 CCACCCCTATTCTACCTACATAATAAACAATAAAACCAACACATAAAAAATTTAACA 851
 QY 961 GTGAGGCTGCCCTGCTGCTGGTCCACCCCTCTCTGGCT 1000
 Db 850 ATAAAACTACCCCTACTATCTAAATCACCCCTCTCTACT 811

RESULT 4
 AAQ20379/C
 ID AAQ20379 standard; cDNA; 549 BP.
 XX AC AAQ20379;
 XX DT
 XX 13-APR-1992 (first entry)
 DE Sequence encoding porcine C-type natriuretic peptide (CNP) contg.
 DE precursor.
 XX KW Natriuretic; hypotensive; hormone; ss.
 XX OS Fig.
 XX FH Key Location/Qualifiers
 FT CDS 363..524
 FT /*tag= a
 XX EP466174-A.
 XX PN
 XX 15-JAN-1992.
 XX PF 12-JUL-1991; 91EP-0111629.
 XX PR 13-JUL-1990; 90JP-0186582.
 XX PA (MATS/) MATSUO H.
 XX PI Matsuo H, Kangawa K, Minamino N;
 XX DR WPI; 1992-017805/03.
 DR P-PSDB; AAR20182.
 XX PT New C-type natriuretic peptide(s) isolated from porcine brain -
 PT possess natriuretic and hypotensive activity
 XX PS Disclosure; Fig 6; 20pp; English.
 XX CC CNP can be isolated from porcine brains using antibodies to CNP-22
 CC (JA105047/90) and sequenced. The 22 and 53 C-terminal AAs have
 CC natriuretic and hypotensive activities, and are claimed (see AAR20182).
 XX QY Sequence 549 BP; 85 A; 212 C; 178 G; 74 T; 0 other;

Query Match 4.5%; Score 44.6; DB 13; Length 549;
 Best Local Similarity 49.5%; Pred. No. 0.021;
 Matches 145; Conservative 0; Mismatches 144; Indels 4; Gaps 1;
 QY 89 TGGCTACCCCTCTGCTCTGCTACCCAGGCCCCCAACCCGCCCCCAGCAG 148
 Db 306 TGTGCGCTTCTTCTGACCGCGCGGCTGGGCTGGGCTCTCTCCCTGGCG 247
 QY 149 CTGTTCTCAGGCTCTCAGC----CTGTCGATTGCTTCTGCTGGCTGGGAGATGAG 204
 Db 246 GAGTTTCGAGGAGCTTCTGCGGCGCTCCGGGCTTGGCTTCGAGGCGGAGAGAGA 187
 QY 205 GTGGGAGAAACACGAGGAGTGGTGGTGGAGTGAAGACAGACGCGGTGGGA 264
 Db 186 GCGTGAGACGACGAGGCGGAGAGTGGAGAGTGCATGCTGGGGTGA 127
 QY 265 GTTCAGGAGAAATCTCTGGGCTGGGAGTGGTGGGATCAACTGTCCCATTTGCTGCA 324
 Db 126 GGGGCGCAGATCGCGGCGAGAGGTGCAGGGGCGGCGGCGGCGGCGGCGGCGG 67
 QY 325 GGCTGTCTTGGGGCAGGGAAGGGGATGGGGGCCCATAGCAGTGTGCTGCTCAGC 377

Db 56 CGGACGACACTGCGCGGCTGCGGCGGCGGAGTCCAGTCTCGCGGC 14
 RESULT 5
 AAQ20467/C
 ID AAQ20467 standard; cDNA; 549 BP.
 XX AC AAQ20467;
 XX DT 14-APR-1992 (first entry)
 XX DE Sequence encoding porcine C-type natriuretic peptide (CNP) precursor.
 XX KW Hypotensive agent; natriuretic; brain-specific promoter; ss.
 XX OS Sus scrofa domestica.
 XX FH Key Location/Qualifiers
 FT exon 144..521
 FT /*tag= a
 XX EP466175-A.
 XX PN
 XX 15-JAN-1992.
 XX PF 12-JUL-1991; 91EP-0111630.
 XX PR 13-JUL-1990; 90JP-0186583.
 XX PA (SUNR) SUNTORY LTD.
 XX PI Matsuo H, Kangawa K, Minamino N, Tanaka S, Fuchimura K;
 PI Tawaragi Y;
 DR WPI; 1992-017806/03.
 DR P-PSDB; AAR20074.
 XX PT DNA encoding porcine C-natriuretic peptide - or precursor and
 PT fragments with natriuretic and hypotensive activities
 XX PS Claim 12; Fig 5; 23pp; English.
 XX CC Genomic and cDNA encoding porcine CNP are claimed. Also claimed are
 CC fragments encoding the last 22 AAs (CNP-22) and the last 53 AAs
 CC (CNP-53). Proteins encoded are also claimed. The full length DNA
 CC includes a promoter which may act in a brain-specific manner so
 CC could be used to control expression of other genes in the brain
 CC (see AAQ20465).
 XX QY Sequence 549 BP; 85 A; 212 C; 178 G; 74 T; 0 other;

Query Match 4.5%; Score 44.6; DB 13; Length 549;
 Best Local Similarity 49.5%; Pred. No. 0.021;
 Matches 145; Conservative 0; Mismatches 144; Indels 4; Gaps 1;
 QY 89 TGGCTACCCCTCTGCTCTGCTACCCAGGCCCCCAACCCGCCCCCAGCAG 148
 Db 306 TGTGCGCTTCTTCTGACCGCGCGGCTGGGCTGGGCTCTCTCCCTGGCG 247
 QY 149 CTGTTCTCAGGCTCTCAGC----CTGTCGATTGCTTCTGCTGGCTGGGAGATGAG 204
 Db 246 GAGTTTCGAGGAGCTTCTGCGGCGCTCCGGGCTTGGCTTCGAGGCGGAGAGAGA 187
 QY 205 GTGGGAGAAACACGAGGAGTGGTGGTGGAGTGAAGACAGACGCGGTGGGA 264
 Db 186 GCGTGAGACGACGAGGCGGAGAGTGGAGAGTGCATGCTGGGGTGA 127
 QY 265 GTTCAGGAGAAATCTCTGGGCTGGGAGTGGTGGGATCAACTGTCCCATTTGCTGCA 324
 Db 126 GGGGCGCAGATCGCGGCGAGAGGTGCAGGGGCGGCGGCGGCGGCGGCGGCGG 67
 QY 325 GGCTGTCTTGGGGCAGGGAAGGGGATGGGGGCCCATAGCAGTGTGCTGCTCAGC 377

Db 187508 GCGCGAAGCGCGCGTGGTATTGGCGGAGTCACTGTTTCAGCGTACTGTCGGGGAG 187567

QY 675 GCCTAGGTAGGATGTGGAAATATGGGAGGAGGAGGAGGATCTGTCTACCTAGGAGGCATC 734

Db 187568 GCCTCGGTCCAAAGTGTGCTGAGGTAGAGGTGGTATCCGCCGGAACCTCGTGGAGTT 187627

QY 735 CTCATCCATCCTTGGCCCGCTGGGACAAGAGA 763

Db 187628 CAAGTTCAACGTGTGACGGCGCGCAGAAGA 187656

RESULT 8

ABAL17086

ID ABAL17086 standard; DNA; 30032 BP.

XX

AC

XX

XX

DT 23-JAN-2002 (first entry)

XX

DE

XX Human nervous system related polynucleotide SEQ ID NO 9417.

XX

KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide; immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary; antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective; antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal; antiparasitic; cardiant; immune disorder; cardiovascular disorder; neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX

OS Homo sapiens.

XX

PN WO200159063-A2.

XX

PD 16-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01334.

XX

31-JAN-2000; 2000US-0179065.

PR

04-FEB-2000; 2000US-0180628.

PR

24-FEB-2000; 2000US-0184664.

PR

02-MAR-2000; 2000US-0186350.

PR

16-MAR-2000; 2000US-0189874.

PR

17-MAR-2000; 2000US-0190076.

PR

18-APR-2000; 2000US-0198123.

PR

19-MAY-2000; 2000US-0205515.

PR

07-JUN-2000; 2000US-0209467.

PR

28-JUN-2000; 2000US-0214886.

PR

30-JUN-2000; 2000US-0215135.

PR

07-JUL-2000; 2000US-0215647.

PR

07-JUL-2000; 2000US-0216880.

PR

11-JUL-2000; 2000US-0217487.

PR

11-JUL-2000; 2000US-0217496.

PR

14-JUL-2000; 2000US-0218290.

PR

26-JUL-2000; 2000US-0220963.

PR

26-JUL-2000; 2000US-0220964.

PR

14-AUG-2000; 2000US-0224518.

PR

14-AUG-2000; 2000US-0224519.

PR

14-AUG-2000; 2000US-0225213.

PR

14-AUG-2000; 2000US-0225214.

PR

14-AUG-2000; 2000US-0225266.

PR

14-AUG-2000; 2000US-0225267.

PR

14-AUG-2000; 2000US-0225268.

PR

14-AUG-2000; 2000US-0225270.

PR

14-AUG-2000; 2000US-0225447.

PR

14-AUG-2000; 2000US-0225757.

PR

14-AUG-2000; 2000US-0225758.

PR

14-AUG-2000; 2000US-0225759.

PR

18-AUG-2000; 2000US-0226279.

PR

22-AUG-2000; 2000US-0226681.

PR

22-AUG-2000; 2000US-0226868.

PR

22-AUG-2000; 2000US-0227182.

PR

23-AUG-2000; 2000US-0227009.

PR

30-AUG-2000; 2000US-0228924.

PR

PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241185.
PR 20-OCT-2000; 2000US-0241186.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.

XX AAK83782; PR 14-SEP-2000; 2000US-0232400.
AC 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
DT 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
DE 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
KW 25-SEP-2000; 2000US-0234997.
XX 25-SEP-2000; 2000US-0234998.
KW 26-SEP-2000; 2000US-0235484.
XX 27-SEP-2000; 2000US-0235834.
OS 27-SEP-2000; 2000US-0235836.
PN 29-SEP-2000; 2000US-0236327.
XX 29-SEP-2000; 2000US-0236367.
PD 29-SEP-2000; 2000US-0236368.
XX 29-SEP-2000; 2000US-0236369.
XX 29-SEP-2000; 2000US-0236370.
XX 02-OCT-2000; 2000US-0236802.
XX 02-OCT-2000; 2000US-0237037.
XX 02-OCT-2000; 2000US-0237038.
XX 02-OCT-2000; 2000US-0237039.
XX 02-OCT-2000; 2000US-0237040.
XX 13-OCT-2000; 2000US-0239935.
XX 13-OCT-2000; 2000US-0239937.
XX 20-OCT-2000; 2000US-0240960.
XX 20-OCT-2000; 2000US-0241221.
XX 20-OCT-2000; 2000US-0241785.
XX 20-OCT-2000; 2000US-0241786.
XX 20-OCT-2000; 2000US-0241787.
XX 20-OCT-2000; 2000US-0241808.
XX 20-OCT-2000; 2000US-0241809.
XX 20-OCT-2000; 2000US-0241826.
XX 01-NOV-2000; 2000US-0244617.
XX 08-NOV-2000; 2000US-0246474.
XX 08-NOV-2000; 2000US-0246475.
XX 08-NOV-2000; 2000US-0246476.
XX 08-NOV-2000; 2000US-0246477.
XX 08-NOV-2000; 2000US-0246478.
XX 08-NOV-2000; 2000US-0246523.
XX 08-NOV-2000; 2000US-0246524.
XX 08-NOV-2000; 2000US-0246525.
XX 08-NOV-2000; 2000US-0246526.
XX 08-NOV-2000; 2000US-0246527.
XX 08-NOV-2000; 2000US-0246528.
XX 08-NOV-2000; 2000US-0246532.
XX 08-NOV-2000; 2000US-0246609.
XX 08-NOV-2000; 2000US-0246610.
XX 08-NOV-2000; 2000US-0246611.
XX 08-NOV-2000; 2000US-0246613.
XX 17-NOV-2000; 2000US-0249207.
XX 17-NOV-2000; 2000US-0249208.
XX 17-NOV-2000; 2000US-0249209.
XX 17-NOV-2000; 2000US-0249210.
XX 17-NOV-2000; 2000US-0249211.
XX 17-NOV-2000; 2000US-0249212.
XX 17-NOV-2000; 2000US-0249213.
XX 17-NOV-2000; 2000US-0249214.
XX 17-NOV-2000; 2000US-0249215.
XX 17-NOV-2000; 2000US-0249216.
XX 17-NOV-2000; 2000US-0249217.
XX 17-NOV-2000; 2000US-0249218.
XX 17-NOV-2000; 2000US-0249244.
XX 17-NOV-2000; 2000US-0249245.
XX 17-NOV-2000; 2000US-0249264.
XX 17-NOV-2000; 2000US-0249265.
XX 17-NOV-2000; 2000US-0249297.
XX 17-NOV-2000; 2000US-0249299.
XX 01-DEC-2000; 2000US-0250300.
XX 01-DEC-2000; 2000US-0250160.
XX 05-DEC-2000; 2000US-0250391.
XX 05-DEC-2000; 2000US-0251030.
XX 05-DEC-2000; 2000US-0251988.
XX 05-DEC-2000; 2000US-0256719.
Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38594.
Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
cytostatic; gene therapy; vaccine; metastasis; ds.
Homo sapiens.
WO200157182-A2.
09-AUG-2001.
17-JAN-2001; 2001WO-US01354.
31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.

PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Disclosure; SEQ ID NO 38594; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK5950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 61710 BP; 11560 A; 18542 C; 18148 G; 13460 T; 0 other;
Query Match 4.2%; Score 42; DB 22; Length 61710;
Best Local Similarity 50.0%; Pred. No. 0.73;
Matches 105; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
QY 195 GGAGATGAGTGGGAGAAACAGCCAGCGGCGAGTGTGTGGAGTGAAGACGACG 254
DB 15325 GGTGATGAGCAGCAGCCAGCGGAGCGCCCTGGAGTTGAGTGGCAGCTGGCAGC 15266
QY 255 GCGTGGGAGGTGAGGAGAGATCTGTGGGCTGGGATGGTGGGCAATCACTGTCC 314
DB 15265 GGGCCCTGGAGGTGAGTACAGCTGGGTATGCCCGGAGGCGCTGGAGGTGAGTGGCAA 15206
QY 315 CATTCGTGAGCTGTCTTGGGCGAGGAGGAGTGGGGCCCATGACAGTGTGGTC 374
DB 15205 TTGGGGGAGGGGCGCTGGAGCGGGGCGAGCCCTGGAGGTGAATAGCAGTGGGTAT 15146
QY 375 AGCCAGGCTGGCTGGGAAGTGGTCCAG 404
DB 15145 GCCCGGAGGCGCTGGAGGTGAGTGGCAG 15116
RESULT 11
ID ABK10687/c
XX ABK10687 standard; DNA; 9359 BP.
AC ABK10687;
XX
DT 05-JUN-2002 (first entry)
XX Transformation vector plasmid pVDH636 DNA.
DE Grass; plant; transgenic; flowering inhibition; inflorescence; gene; ds;
XX

KW tiller production; delayed heading; gibberellic acid; phytohormone; golf;
KW genetically modified grass; athletic field; sport; baseball; cricket;
KW football; rugby; soccer; tennis; lawn; landscaping; cattle; horse; sheep;
KW goat; animal feedstuff; Arabidopsis thaliana; Ath1; homeobox gene;
XX circular; cyclic; pVDH636.
OS Synthetic.
XX

PH Key Location/Qualifiers
FT CDS 4930..6351
FT /*tag= a
FT /product= "Arabidopsis thaliana Ath1"

XX WO200214486-A2.

XX 21-FEB-2002.

XX 16-AUG-2001; 2001WO-EP09570.

XX 18-AUG-2000; 2000US-226422P.

XX 27-NOV-2000; 2000US-253274P.

XX 22-JUN-2001; 2001US-300220P.

XX (ADVA-) ADVANTA SEEDS BV.

XX Van Der Valk P, Van Dun CMP, Smeekens SCM, Proveniers MCG;
XX WPI; 2002-257603/30.
XX P-PSDB; ANU76885.

XX New genetically modified grass useful for growing and/or propagating
PT grass in athletic fields (for sports such as baseball, cricket,
PT football, golf, rugby, soccer and tennis), or as animal feedstuff for
PT cattle, goat, horse and sheep
XX
XX Example 1; Fig 6; 45pp: English.

CC The invention relates to a grass plant which has been genetically
CC modified to substantially inhibit generative propagation. The genetic
CC modification may result in a heritable change in one or more plant
CC characteristics such as inhibition of flowering (or substantial delay
CC that amounts to inhibition), absence of inflorescence, increased
CC production of tillers, delayed heading and inhibition of the
CC developmental switch from vegetative to generative growth. A method of
CC making a grass involves transforming the grass with a nucleic acid which
CC interferes with metabolism of gibberellic acid. A grass can be treated by
CC applying a phytohormone to at least partially relieve or reverse a change
CC in plant characteristic resulting from genetic modification. The
CC genetically modified grass is useful for growing and/or propagating grass
CC in athletic fields (for sports such as baseball, cricket, football, golf,
CC rugby, soccer and tennis), lawns, parks and other types of landscaping.
CC The grass is also useful as an animal feedstuff for cattle, goats, horses
CC and sheep, due to its increased vegetative growth, improved digestibility
CC and/or nutritional value as animal feedstuff. This sequence represents
CC plasmid pVDH636 DNA. This plasmid was used as a transformation vector for
CC production of transgenic grasses expressing the Arabidopsis thaliana
CC homeobox gene, Ath1.

XX Sequence 9359 BP; 2377 A; 2222 C; 2133 G; 2624 T; 3 other;

Query Match 4.1%; Score 41.4; DB 24; Length 9359;
Best Local Similarity 50.6%; Pred. No. 0.51;
Matches 119; Conservative 3; Mismatches 112; Indels 1; Gaps 1;

QY 119 CCAGGGCCCCCAACCCGCCGCCCGCCAGCAGCTGTTCTCAGCCCTCTCAGCTGTCTGATT 178
DB 7416 CCGGCCCCCAGCAGCCGCCGCCCGCCAGCAGCTGTTCTCTTTCTTTCTTTCTTTCTTTT 7357

QY 179 TGCTTGTCTGCTGGGAGAAATGAGTGGGAGAAACAGCCAGCCAGCCAGCTGGTGTG 238
DB 7356 TTTTCTTTTGTGCGAGACGTCGCGTGCCTGCGGCAACTCAGGTGATAGTGGGGG 7297

Qy 239 GACTGAAGACAGACGCGGTGGGAGTCTCAGAGAGAAATCTGCTGGCTGGGGATGCTG 298
Db 7296 GHVTTGGAGACTATTGTCAGTGTGCTGGACTGGGTGGGTGGGTGGGTGGGTGGGC 7237
Qy 299 TGGGCATCACTGCCATTGCT-GCAGGCTGGTCTTGGGCGAGGAGGAGGATG 352
Db 7236 TGGGTTGCTATGATCGTGATAGCACTTTGGGCTTTAGGAACCTTAGGGGTTG 7182

RESULT 12
ABK11039/c
ID ABK11039 standard; DNA; 9359 BP.
XX AC ABK11039;
XX AC ABK11039;
DT DT
XX 05-JUN-2002 (first entry)
XX pVDH636 vector containing Arabidopsis Ath1 gene.
DE pVDH636 vector; circular; cyclic; grass; plant; herbicide resistance;
XX baseball; cricket; football; golf; rugby; soccer; tennis; lawn; park;
KW athletic field; animal feedstuff; grass flowering inhibitor;
KW Ath1; gene; ds.
XX Arabidopsis thaliana.
OS Zea mays.
OS Agrobacterium tumefaciens.
OS Oryza sativa.
OS Escherichia coli.
OS Cauliflower mosaic virus.
OS Synthetic.
XX Key Location/Qualifiers
FH CDS 4930..6360
FT /*tag= a
FT /product= "Ath1"
FT /partial
FT /note= "No stop codon"
XX WO200214524-A2.
XX 21-FEB-2002.
XX 16-AUG-2001; 2001WO-EP09572.
XX 18-AUG-2000; 2000US-226422P.
PR 27-NOV-2000; 2000US-253274P.
PR 27-NOV-2000; 2000US-253327P.
PR 22-JUN-2001; 2001US-300220P.
XX (ADVA-) ADVANTA SEEDS BV.
XX Van Der Valk P, Van Dun CMP, Smeekens SCM, Proveniers MCG;
PI WPI; 2002-257611/30.
DR P-PSDB; AAU76514.
XX New genetically modified grasses that express inhibited generative
PT propagation, or herbicide resistance, useful for forage (e.g. cattle
PT feedstuff) or amenity purposes (e.g. for use in an athletic field, lawn
PT or park)
XX Example 1; Fig 6; 56pp; English.
PS The invention relates to a grass plant, which has been genetically
XX modified to substantially inhibit generative propagation and carry
CC herbicide resistance. The grass is useful for growth and/or propagation
CC of grasses. The grass is used to play at least one sport (e.g. baseball,
CC cricket, football, golf, rugby, soccer or tennis), or used at least in a
CC portion of an athletic field, lawn or park. The grass is fed to animal
CC (e.g. cattle, goat, horse or sheep) or used as an animal feedstuff.
CC The present sequence represents the coding sequence of pVDH636 vector
CC containing the Arabidopsis Ath1 gene, which was used to inhibit flowering

CC in grasses.
XX Sequence 9359 BP; 2377 A; 2222 C; 2133 G; 2624 T; 3 other;
SQ Best Local Similarity 4.1%; Score 41.4; DB 24; Length 9359;
Best Local Similarity 50.6%; Pred. No. 0.51;
Matches 119; Conservative 3; Mismatches 112; Indels 1; Gaps 1;
Qy 119 CCAGGCCCCCACCACCCGCCCCCCCCCAGCAGCTGTTCTCAGGCTCTCAGCCTGCTGAT 178
Db 7416 CCGGCCCCCAGCAGCCCCCAGCAGCTGTTCTTTTCTTTTCTTTTCTTTTCTTTT 7357
Qy 179 TGCTTGTCTGGCTGGGAGAAATGAGTGGGAGAAAACCCAGCCAGGGCAGTTGGTGTG 238
Db 7356 TTTTCTTTTGGCTGCGAGACGTGGGTGCGTGGCGACACACTCAGGTCATAGTGGGGG 7297
Qy 239 GAGTGAAGACAGACGCGCGGTGGGAGGTCAGGAGAGAAATCTGCTGGGCTGGGATGGTG 298
Db 7296 GHVTTGGAGACTATTGTCAGTGTGCTGGACTGGGTGGGTGGGTGGGTGGGTGGGC 7237
Qy 299 TGGGCATCACTGCCATTGCT-GCAGGCTGGTCTTGGGCGAGGAGGAGGATG 352
Db 7236 TGGGTTGCTATGATCGTGATAGCACTTTGGGCTTTAGGAACCTTAGGGGTTG 7182

RESULT 13
AAQ20466/c
ID AAQ20466 standard; DNA; 1894 BP.
XX AC AAQ20466;
XX AC AAQ20466;
DT DT
XX 14-APR-1992 (first entry)
XX Sequence of chromosomal gene, including promoter, for porcine
DE C-type natriuretic peptide (CNP) precursor.
XX Hypotensive agent; natriuretic; brain-specific promoter; ss.
XX Sus scrofa domestica.
XX Key Location/Qualifiers
FH Promoter 80..84
FT /*tag= a
FT /label= Y box
FT promoter 88..93
FT /*tag= b
FT /label= GC box
FT promoter 100..105
FT /*tag= c
FT /note= "GC box"
FT TATA_signal 133..138
FT /*tag= d
FT exon 310..399
FT /*tag= e
FT exon 839..1229
FT /*tag= f
XX EP466175-A.
PN PN
XX 15-JAN-1992.
PD PD
XX 12-JUL-1991; 91EP-0111630.
PF PF
XX 13-JUL-1990; 90JP-0186583.
PR PR
XX (SUNR) SUNTORY LTD.
PA PA
XX Matsuo H, Kangawa K, Minamino N, Tanaka S, Fuchimura K;
PI Tawaragi Y;
XX WPI; 1992-017806/03.
DR P-PSDB; AAR20074.
XX

```
PT DNA encoding porcine C-natriuretic peptide - or precursor and
PT fragments with natriuretic and hypotensive activities
XX
PS Claim 13; Fig 3; 23pp; English.
XX
CC Genomic and cDNA encoding porcine CPN are claimed. Also claimed are
CC fragments encoding the last 22 AAs (CPN-22) and the last 53 AAs
CC (CPN-53). Proteins encoded are also claimed. The full length DNA
CC includes a promoter which may act in a brain-specific manner so
CC could be used to control expression of other genes in the brain
CC (see AAQ20465).
XX
SQ Sequence 1894 BP; 339 A; 614 C; 639 G; 302 T; 0 other;

Query Match 4.0%; Score 39.6; DB 13; Length 1894;
Best Local Similarity 45.9%; Pred. No. 0.87;
Matches 135; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 175 GATTGCTTCTGCTGGCTGGGAGAAATGAGTGGGAGAAACCAACAGGCGGCGAGTTGGT 234
DB 175 GATTGCTTCTGCTGGCTGGGAGAAATGAGTGGGAGAAACCAACAGGCGGCGAGTTGGT 234
QY 235 GTTGGAGTGAAGACAGACGCGGTGGGAGGTCAGGAGAGAAATCTGCTGGCTGGGGAT 294
DB 235 GTTGGAGTGAAGACAGACGCGGTGGGAGGTCAGGAGAGAAATCTGCTGGCTGGGGAT 294
QY 322 GCGAGAGGTGCATGGTGCCTGTGGGTGGAGGGCGGAGATCGGCGGCGAGAGGTGCAC 263
DB 322 GCGAGAGGTGCATGGTGCCTGTGGGTGGAGGGCGGAGATCGGCGGCGAGAGGTGCAC 263
QY 295 GGTGTGGGCATCAACTGCTCCATGCTGCAGGCTGCTTTGGGCGAGGGAAGGGATGGG 354
DB 295 GGTGTGGGCATCAACTGCTCCATGCTGCAGGCTGCTTTGGGCGAGGGAAGGGATGGG 354
QY 262 GGGCGCGGCGGCGAGAGTGGCAAGCGAGCGCGGAGAGAGTGGCGGCGTGGGTGGGG 203
DB 262 GGGCGCGGCGGCGAGAGTGGCAAGCGAGCGCGGAGAGAGTGGCGGCGTGGGTGGGG 203
QY 355 GGGCCATAGCAGTGCCTGGGAGAAATGAGTGGGAGAAACCAACAGGCGGCGAGTTGGT 414
DB 355 GGGCCATAGCAGTGCCTGGGAGAAATGAGTGGGAGAAACCAACAGGCGGCGAGTTGGT 414
QY 202 CGGAGTCCAGTGCCTGGGCGGCGGCGGAGTGGGTGGGTCTGAGCGGCTGGCTCGG 143
DB 202 CGGAGTCCAGTGCCTGGGCGGCGGCGGAGTGGGTGGGTCTGAGCGGCTGGCTCGG 143
QY 415 GAGCCAGGAAGCCCTCCCAAGGTTGTTGGCTAGTTCCTCTGTCATCAGCGGCC 468
DB 415 GAGCCAGGAAGCCCTCCCAAGGTTGTTGGCTAGTTCCTCTGTCATCAGCGGCC 468
QY 142 CGCCTTTAATAATCAACCTGCCGTGATGTCATCTCCCGCCCAAGGCGGCC 89
DB 142 CGCCTTTAATAATCAACCTGCCGTGATGTCATCTCCCGCCCAAGGCGGCC 89

RESULT 15
ID AAT71220/c
AC AAT71220 standard; DNA; 1894 BP.
XX
AC AAT71220;
XX
DT 30-SEP-1997 (first entry)
XX
DE Porcine C-type natriuretic peptide gene.
XX
KW C-type natriuretic peptide; CNP-22; CNP-53; hypotensive; ss.
XX
OS Sus scrofa.
XX
FH Key Location/Qualifiers
FT misc_signal 80..84 /tag= a
FT misc_signal 88..93 /label= Y-box
FT misc_signal 100..105 /tag= b
FT misc_signal 100..105 /label= GC-box
FT TATA_signal 133..138 /tag= c
FT exon 310..399 /label= GC-boc
FT intron 400..838 /tag= d
FT exon 839..1129 /tag= e
FT primer_bind 968..996 /tag= f
FT /tag= g
FT complement (968..996) /tag= h
```

```
FT primer_bind /note= "primer KF225"
FT 1088..1129 /tag= 1
FT /note= "primer KF226"
XX
XX EP771875-A1.
XX
XX 07-MAY-1997.
XX
XX 12-JUL-1991; 91EP-0111630.
XX
XX 13-JUL-1990; 90JP-0186583.
XX
XX (MATSUO) MATSUO H.
XX (SUNR ) SUNTORY LTD.
XX
XX Kangawa K, Kawashima K, Matsuo H, Minamino N, Tanaka S;
XX Tawaragi Y;
XX
XX WPI; 1997-247417/23.
XX P-PSDB; AAW21990.
XX
XX New isolated porcine C-type natriuretic peptide gene - used to study
XX the mechanisms of biosynthesis and physiological actions of CNP and
XX to develop pharmaceutical applications
XX
XX Example 1; Fig 3a-b; 18pp; English.
XX
XX A chromosomal DNA fragment (AAT71220) includes the promoter region
XX and structural gene coding for the precursor protein (AAW21990) of
XX porcine C-type natriuretic peptides CNP-53 (AAW21991) and CNP-22
XX (AAW21992). It was isolated from a porcine chromosomal gene library
XX using a probe prep. by PCR amplification of porcine DNA using
XX primers (see also AAT71222 and AAT71224) based on the N- and C-terminal
XX portions of CNP-53. A signal peptide is present in the N-terminal
XX region of the precursor protein (propro-CNP) and both CNP-22 and
XX CNP-53 are secreted when the structural gene is expressed in COS-1
XX cells. The structural gene can be used to produce CNP polypeptides
XX for use in studying the biosynthesis and physiological actions of
XX CNP and to develop pharmaceutical applications. The promoter
XX region may be useful as a brain-specific promoter.
XX
XX Sequence 1894 BP; 339 A; 615 C; 638 G; 302 T; 0 other;

Query Match 4.0%; Score 39.6; DB 18; Length 1894;
Best Local Similarity 45.9%; Pred. No. 0.87;
Matches 135; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 175 GATTGCTTCTGCTGGCTGGGAGAAATGAGTGGGAGAAACCAACAGGCGGCGAGTTGGT 234
DB 175 GATTGCTTCTGCTGGCTGGGAGAAATGAGTGGGAGAAACCAACAGGCGGCGAGTTGGT 234
QY 382 GCTTGGCTTCGGAGGCGGCGGAGAGAGAGAGCGTGGAGAGAGAGCGGCGGCGAGCGCT 323
DB 382 GCTTGGCTTCGGAGGCGGCGGAGAGAGAGAGAGCGTGGAGAGAGAGCGGCGGCGAGCGCT 323
QY 235 GTTGGAGTGAAGACAGACGCGGTGGGAGGTCAGGAGAGAAATCTGCTGGCTGGGGAT 294
DB 235 GTTGGAGTGAAGACAGACGCGGTGGGAGGTCAGGAGAGAAATCTGCTGGCTGGGGAT 294
QY 322 GGGAGAGGTGCATGGTGCCTGTGGGAGGAGGCGGCGGCGAGAGAGGTGCAC 263
DB 322 GGGAGAGGTGCATGGTGCCTGTGGGAGGAGGCGGCGGCGAGAGAGGTGCAC 263
QY 295 GGTGTGGGCATCAACTGCTCCCATGCTGCAGGCTGGTCTTGGGCGAGGGAAGGGATGGG 354
DB 295 GGTGTGGGCATCAACTGCTCCCATGCTGCAGGCTGGTCTTGGGCGAGGGAAGGGATGGG 354
QY 262 GGGCGGCGGCGGCGAGAGTGGCAAGCGAGCGGCGGAGAGAGTGGCGGCGTGGGTGGGG 203
DB 262 GGGCGGCGGCGGCGAGAGTGGCAAGCGAGCGGCGGAGAGAGTGGCGGCGTGGGTGGGG 203
QY 355 GGGCCATAGCAGTGCCTGGGAGAAATGAGTGGGAGAAACCAACAGGCGGCGAGTTGGT 414
DB 355 GGGCCATAGCAGTGCCTGGGAGAAATGAGTGGGAGAAACCAACAGGCGGCGAGTTGGT 414
QY 202 CGGAGTCCAGTGCCTGGGCGGCGGCGGAGTGGGTGGGTCTGAGCGGCTGGCTCGGCTCG 143
DB 202 CGGAGTCCAGTGCCTGGGCGGCGGCGGAGTGGGTGGGTCTGAGCGGCTGGCTCGGCTCG 143
QY 415 GAGCCAGGAAGCCCTCCCAAGGTTGTTGGCTAGTTCCTCTGTCATCAGCGGCC 468
DB 415 GAGCCAGGAAGCCCTCCCAAGGTTGTTGGCTAGTTCCTCTGTCATCAGCGGCC 468
QY 142 CGCCTTTAATAATCAACCTGCCGTGATGTCATCTCCCGCCCAAGGCGGCC 89
DB 142 CGCCTTTAATAATCAACCTGCCGTGATGTCATCTCCCGCCCAAGGCGGCC 89

RESULT 15
AAS29976/c
ID AAS29976 standard; DNA; 5786 BP.
XX
```

AC AAS29976;
XX
XX DT 21-NOV-2001 (first entry)
XX
XX DE Human lung antigen genomic DNA #46.
XX
XX
KW Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-infertility; food additive.
XX
XX Homo sapiens.
XX
XX WO200155303-A2.
XX
XX PN
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01301.
XX
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232968.
PR 14-SEP-2000; 2000US-0232997.
PR 14-SEP-2000; 2000US-0232998.
PR 14-SEP-2000; 2000US-0232999.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.


```
Db 194 LFLGQTGLTTEEMKLALEPGENFCMGGPGVIMSRVLRMPVPHIGKCLREMYTHDEV 253
QY 522 AGCCCTCGGTGTTCCCATGCTCTCTCTGTCGCCCAATTTTAGGCCCATGTGATTG 581
Db 254 EVGRCVRFAGVQCVMSYEMRQLFYENYQNKKGVIKRLHNSKIHQAITLHPNKNPPYQY 313
QY 582 GGGCTATGTGACTCATGCTCTAAGTGCTTGGGCCAGGAGCTGTGGCCACCTTTAAATG 641
Db 314 RLHMYMLSRKISELRHRIQIHLREIVLMSKYSNTEIHKEDLQLGIPPFMRFPQOREI 373
QY 642 CCAGCCAGTCATGTCGCCGAGTTTGGGTTAGGCTAGGTAGGATTTGGAATATGGGA 701
Db 374 LEWEFLTKYLSAVDGOOPRRGMDSAQREALDDIVQMVMENANAKTRIGRIIDFKEIQ 433
QY 702 GGAGCGAGGATCTGTCTACCTAGGAGGCATCCTCATCCATCCCTGTCGCCCTGGACAGA 761
Db 434 YGYRRVPMWGAEYILDLLLYKKHKKMTVPVRRHAYLQQTFSKIQFVEHEELDAOEL 493
QY 762 GAACCTTGAACCTTGGTAGGGCCTCAGGACGATGCTGCGTGGCCCTTGGGAATCTGGA 821
Db 494 AKRINQESGSLFSLNSLKLVPQLPGSKSEHKPKDKKINILPLSGRDMFVRPMGN 553
QY 822 TTGTCCTGGTCATAGTTCTTATCTGTGACCAACACCCCTTAGCTGCCAGGCTTTGGACA 881
Db 554 FEKTLIPNQWKLVLVLLFNSDSNPDKAKQVELMRDYRIKYPKADMQILPVSGEFSRALA 613
QY 882 TGG 884
Db 614 LEV 616
```

RESULT 2

US-10-079-854-240/c
; Sequence 240, Application US/10079854
; Publication No. US20030034368A1

GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAL21C1
; CURRENT APPLICATION NUMBER: US/10/079,854

; CURRENT FILING DATE: 2002-02-22

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 428

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 240

; LENGTH: 5786

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-10-079-854-240

Query Match 3.9%; Score 39.2; DB 9; Length 5786;
Best Local Similarity 51.1%; Pred. No. 0.029;
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
QY 29 TTGCTATTGTTACTGCTTTTACGTCCTTGGAAAAAGTTAGCACAAAGGGCTGCTTG 88
Db 4527 TTTTCTTGGATAGAGTCTCAGTCTGTACCCAGGTGTGAGTACAGTGGGTGATCTC 4468
QY 89 TGGCTACACCCCTCTGCTCTGCTGCTCAGCCACCCAGGCCCCCAACCCGCCCCCAGCAG 148
Db 4467 AGCTCCCTGCTCTGCTGCTCCAGGTTCAACTGATTTCTCCACCTCAGCCCTCCAGAGCAG 4408
QY 149 CTGTTCTCAGGCTCTCAGCTGTCTGATTTGCTGTGCTGCTGGGAGATGAGGTGG 208
Db 4407 CTGGGACCACGAGCATGCACCACTATGCCAGCTAATTTTGGTGGGTATTTTGGTAG 4348
```

RESULT 3

US-09-764-878-240/c

; Sequence 240, Application US/09764878
; Patent No. US20020090615A1

GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAL21
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 240
; LENGTH: 5786
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-878-240

Query Match 3.9%; Score 39.2; DB 10; Length 5786;
Best Local Similarity 51.1%; Pred. No. 0.029;
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
QY 29 TTGCTATTGTTACTGCTTTTACGTCCTTGGAAAAAGTTAGCACAAAGGGCTGCTTG 88
Db 4527 TTTTCTTGGATAGAGTCTCAGTCTGTACCCAGGTGTGAGTACAGTGGGTGATCTC 4468
QY 89 TGGCTACACCCCTCTGCTCTGCTGCTCAGCCACCCAGGCCCCCAACCCGCCCCCAGCAG 148
Db 4467 AGCTCCCTGCTCTGCTGCTCCAGGTTCAACTGATTTCTCCACCTCAGCCCTCCAGAGCAG 4408
QY 149 CTGTTCTCAGGCTCTCAGCTGTCTGATTTGCTGTGCTGCTGGGAGATGAGGTGG 208
Db 4407 CTGGGACCACGAGCATGCACCACTATGCCAGCTAATTTTGGTGGGTATTTTGGTAG 4348
```

RESULT 4

US-09-989-442-47/c
; Sequence 47, Application US/09989442
; Publication No. US20030013649A1

GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: P2J08

; CURRENT APPLICATION NUMBER: US/09/989,442

; CURRENT FILING DATE: 2001-11-21

; Prior Application Number: 60/179,065

; Prior Filing Date: 2000-01-31

; Prior Application Number: 60/180,628

; Prior Filing Date: 2000-02-04

; Prior Application Number: 60/214,886

; Prior Filing Date: 2000-06-28

; Prior Application Number: 60/217,487

; Prior Filing Date: 2000-07-11

; Prior Application Number: 60/225,758

; Prior Filing Date: 2000-08-14

; Prior Application Number: 60/220,963

; Prior Filing Date: 2000-07-26

; Prior Application Number: 60/217,496

; Prior Filing Date: 2000-07-11

; Prior Application Number: 60/225,447

; Prior Filing Date: 2000-08-14

; Prior Application Number: 60/218,290

; Prior Filing Date: 2000-07-14

; Prior Application Number: 60/225,757

; Prior Filing Date: 2000-08-14

; Prior Application Number: 60/226,868

; Prior Filing Date: 2000-08-22

; Prior Application Number: 60/216,647

; Prior Filing Date: 2000-07-07

; Prior Application Number: 60/225,267

; Prior Filing Date: 2000-08-14

; Prior Application Number: 60/216,880

; Prior Filing Date: 2000-07-07

; Prior Application Number: 60/225,270

; Prior Filing Date: 2000-08-14

; Prior Application Number: 60/251,869

; Prior Filing Date: 2000-12-08

; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/241,809
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/249,299
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/236,327
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474

; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:06:43 ; Search time 1290.57 Seconds
(without alignments)
12561.622 Million cell updates/sec

Title: US-09-691-220-3_COPY_5000_6000

Perfect score: 1001

Sequence: 1 tgatctccctcccccacc.....tgggtcaccctctctggtg 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: gb_gss:**

18: em_gss_hum:**

19: em_gss_inv:**

20: em_gss_pin:**

21: em_gss_vrt:**

22: em_gss_fun:**

23: em_gss_mam:**

24: em_gss_mus:**

25: em_gss_other:**

26: em_gss_pro:**

27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	74	7.4	1101	17	CNS000397
c 2	50	5.0	974	17	BH157536
c 3	49.8	5.0	620	17	AG100622
c 4	49.4	4.9	1046	14	BQ682356
c 5	48.4	4.8	1398	13	BM479595
c 6	47.8	4.8	817	17	CNS041N8
c 1	74	7.4	1101	17	CNS000397
c 2	50	5.0	974	17	BH157536
c 3	49.8	5.0	620	17	AG100622
c 4	49.4	4.9	1046	14	BQ682356
c 5	48.4	4.8	1398	13	BM479595
c 6	47.8	4.8	817	17	CNS041N8

7	47.8	4.8	910	17	BH162948
8	47.8	4.8	974	17	BH157536
9	47.8	4.8	997	17	CNS006DN
10	46.4	4.6	384	17	CNS035G7
11	46.4	4.6	1101	17	CNS0172R
12	46.2	4.6	936	17	CNS01608
13	46	4.6	884	17	CNS00600
14	45.8	4.6	951	9	AL543262
15	45.6	4.6	758	17	AG080593
16	45.6	4.6	895	17	A2667655
17	45.6	4.6	1101	17	CNS0020T
18	45.4	4.5	1249	12	BE873655
19	45.4	4.5	1417	14	BM810236
20	45.2	4.5	897	17	A2186654
21	45.2	4.5	1040	17	CNS01671
22	45	4.5	544	17	CNS015XA
23	45	4.5	706	17	AG062296
24	45	4.5	724	17	A2197686
25	45	4.5	730	17	AG076255
26	45	4.5	1113	17	AG030979
27	44.8	4.5	324	17	CNS001CM
28	44.8	4.5	894	17	AG081742
29	44.8	4.5	925	17	CNS0091P
30	44.8	4.5	962	17	AG152777
31	44.6	4.5	506	9	AL549672
32	44.6	4.5	857	17	BH157466
33	44.4	4.4	1047	17	CNS006KX
34	44.4	4.4	1282	14	BQ65718
35	44.2	4.4	556	17	CNS046C5
36	44.2	4.4	777	17	AG104642
37	44.2	4.4	923	9	AL543393
38	44.2	4.4	925	17	CNS0091P
39	44.2	4.4	1041	17	CNS00LFU
40	44.2	4.4	1201	9	AL581589
41	44	4.4	461	17	BH750315
42	44	4.4	694	17	AG100694
43	44	4.4	895	17	CNS06LLW
44	44	4.4	996	17	CNS040W3
45	44	4.4	1101	17	CNS00K66

ALIGNMENTS

RESULT 1	CNS000397	1101 bp	DNA	linear	GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC #				
DEFINITION	BACR08A10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL063912				
VERSION	AL063912.1	GI:4941769			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	Genoscope.				
AUTHORS	Direct Submission				
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :				
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)				
COMMENT	<p>Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoss in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial</p>				

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

FAX: 301 838 3543
Email: bjloftus@tigr.org

High quality sequence stop: 206.

Location/Qualifiers
1 074

```
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/close lib="Entamoeba histolytica Sheared DNA"
```

/note=Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith, Y. C. and Venter, J. C. (1987). The λ 1074 method used for

A.O. and Venter, J.C. (making small insert libraries for whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999).

212 c 291 g 233 t

509
608
707
806
905
1004
1103
1202
1301
1400
1500

ty	5.0%;	Score 50;	DB 17;	Length 9/4;
ty	49.2%;	Pred. No. 0.016;		

Derivative 0; Mismatches 135; Indels 0; Gaps 0;

TTGGCTCCTGGCCTCACCACAGGCCCCCAACCCCGCCCCCAGCAGTG 151

766

CTCAGCCTGTCTGATTTGCTTGTCTGGCCTGGGGAGATGAGTGGGAG 211

CCCCCCCCCCCCCCCCCCCCCCCCACAGGGGGGGGGGGGGGG 826

AGGGCAGTTCGTCTTGGAGTGACAGACGACGGCGCTGGCCACGCTCAGG 271

[illegible]

886

TGGGCTGGGGATGGTGTGGGCATCAACTGTCCCCATTGCTGCAGGCTGGT 331

946

GAAGGGGATGGGGG 357

GGGGGGGGGGGGG 972

Accession	Length	Model	Score	GC	GC3	GC3+4	GC3+5	GC3+6	GC3+7	GC3+8	GC3+9	GC3+10	GC3+11	GC3+12	GC3+13	GC3+14	GC3+15	GC3+16	GC3+17	GC3+18	GC3+19	GC3+20	GC3+21	GC3+22	GC3+23	GC3+24	GC3+25	GC3+26	GC3+27	GC3+28	GC3+29	GC3+30	GC3+31	GC3+32	GC3+33	GC3+34	GC3+35	GC3+36	GC3+37	GC3+38	GC3+39	GC3+40	GC3+41	GC3+42	GC3+43	GC3+44	GC3+45	GC3+46	GC3+47	GC3+48	GC3+49	GC3+50	GC3+51	GC3+52	GC3+53	GC3+54	GC3+55	GC3+56	GC3+57	GC3+58	GC3+59	GC3+60	GC3+61	GC3+62	GC3+63	GC3+64	GC3+65	GC3+66	GC3+67	GC3+68	GC3+69	GC3+70	GC3+71	GC3+72	GC3+73	GC3+74	GC3+75	GC3+76	GC3+77	GC3+78	GC3+79	GC3+80	GC3+81	GC3+82	GC3+83	GC3+84	GC3+85	GC3+86	GC3+87	GC3+88	GC3+89	GC3+90	GC3+91	GC3+92	GC3+93	GC3+94	GC3+95	GC3+96	GC3+97	GC3+98	GC3+99	GC3+100	GC3+101	GC3+102	GC3+103	GC3+104	GC3+105	GC3+106	GC3+107	GC3+108	GC3+109	GC3+110	GC3+111	GC3+112	GC3+113	GC3+114	GC3+115	GC3+116	GC3+117	GC3+118	GC3+119	GC3+120	GC3+121	GC3+122	GC3+123	GC3+124	GC3+125	GC3+126	GC3+127	GC3+128	GC3+129	GC3+130	GC3+131	GC3+132	GC3+133	GC3+134	GC3+135	GC3+136	GC3+137	GC3+138	GC3+139	GC3+140	GC3+141	GC3+142	GC3+143	GC3+144	GC3+145	GC3+146	GC3+147	GC3+148	GC3+149	GC3+150	GC3+151	GC3+152	GC3+153	GC3+154	GC3+155	GC3+156	GC3+157	GC3+158	GC3+159	GC3+160	GC3+161	GC3+162	GC3+163	GC3+164	GC3+165	GC3+166	GC3+167	GC3+168	GC3+169	GC3+170	GC3+171	GC3+172	GC3+173	GC3+174	GC3+175	GC3+176	GC3+177	GC3+178	GC3+179	GC3+180	GC3+181	GC3+182	GC3+183	GC3+184	GC3+185	GC3+186	GC3+187	GC3+188	GC3+189	GC3+190	GC3+191	GC3+192	GC3+193	GC3+194	GC3+195	GC3+196	GC3+197	GC3+198	GC3+199	GC3+200	GC3+201	GC3+202	GC3+203	GC3+204	GC3+205	GC3+206	GC3+207	GC3+208	GC3+209	GC3+210	GC3+211	GC3+212	GC3+213	GC3+214	GC3+215	GC3+216	GC3+217	GC3+218	GC3+219	GC3+220	GC3+221	GC3+222	GC3+223	GC3+224	GC3+225	GC3+226	GC3+227	GC3+228	GC3+229	GC3+230	GC3+231	GC3+232	GC3+233	GC3+234	GC3+235	GC3+236	GC3+237	GC3+238	GC3+239	GC3+240	GC3+241	GC3+242	GC3+243	GC3+244	GC3+245	GC3+246	GC3+247	GC3+248	GC3+249	GC3+250	GC3+251	GC3+252	GC3+253	GC3+254	GC3+255	GC3+256	GC3+257	GC3+258	GC3+259	GC3+260	GC3+261	GC3+262	GC3+263	GC3+264	GC3+265	GC3+266	GC3+267	GC3+268	GC3+269	GC3+270	GC3+271	GC3+272	GC3+273	GC3+274	GC3+275	GC3+276	GC3+277	GC3+278	GC3+279	GC3+280	GC3+281	GC3+282	GC3+283	GC3+284	GC3+285	GC3+286	GC3+287	GC3+288	GC3+289	GC3+290	GC3+291	GC3+292	GC3+293	GC3+294	GC3+295	GC3+296	GC3+297	GC3+298	GC3+299	GC3+300	GC3+301	GC3+302	GC3+303	GC3+304	GC3+305	GC3+306	GC3+307	GC3+308	GC3+309	GC3+310	GC3+311	GC3+312	GC3+313	GC3+314	GC3+315	GC3+316	GC3+317	GC3+318	GC3+319	GC3+320	GC3+321	GC3+322	GC3+323	GC3+324	GC3+325	GC3+326	GC3+327	GC3+328	GC3+329	GC3+330	GC3+331	GC3+332	GC3+333	GC3+334	GC3+335	GC3+336	GC3+337	GC3+338	GC3+339	GC3+340	GC3+341	GC3+342	GC3+343	GC3+344	GC3+345	GC3+346	GC3+347	GC3+348	
-----------	--------	-------	-------	----	-----	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	--

ocytes DNA, clone: PTB-103E03.R, genomic survey sequence.

1 GI:16721139

odytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
rv clone:PTB-103F03 B

odytes

; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eutheria; Primates; Catarrhini; Hominidae; Pan.

AUTHORS
 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
 Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE
 BAC end sequences of Library PTB
JOURNAL
 Unpublished
REFERENCE
 2 (bases 1 to 620)
AUTHORS
 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
 Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE
 Direct Submission
JOURNAL
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Chuohiro-cho,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:shimbeg@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
 Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT
 Clones are derived from the chimpanzee BAC library PTB. This BAC end
 was generated during the R&D process and may have higher chance of
 clone tracking errors.
PRIMERS
 Sequencing: M13Rev
LIBRARY
 Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI
FEATURES
 Location/Qualifiers
 1..620
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clones="PTB-103E03.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"
 BASE COUNT 60 a 365 c 97 g 96 t 2 others
 ORIGIN
 Query Match 5.0%; Score 49.8; DB 17; Length 620;
 Best Local Similarity 52.1%; Pred. No. 0.016;
 Matches 111; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
 QY 193 GGGAGAGTGTGGGAGAGAAACAGCCAGGCGAGTGGTGGTGGAGTGAAGAGCAGA 252
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 508 GGGGGGAGGAGGGGGGAGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGAG 449
 QY 253 CGGCGGTGGGAGGTCAGGAGAGATCTGCTGGGCTGGGATGGTGGGCATCACTCT 312
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 448 GGAAGGG 389
 QY 313 CCCATTGCTGAGGCTGCTCTTGGGGCAGGAAGGGGATGGGGCCCATAGCAGTCTGG 372
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 388 GGG 329
 QY 373 TCAGCCAGGCTGGCTGGGAAGTGGTGGCCAGG 405
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 328 CGGCGCGGGAGGCAAGGGGAGTCGGGCATTGG 296

RESULT 4
 BQ682356/c 1046 bp mRNA linear EST 15-JUL-2002
 LOCUS
 AGENCOURT 8212489 NIH_MGC_112 Homo sapiens CDNA clone IMAGE:6260920
 5', mRNA sequence.
ACCESSION
 BQ682356
VERSION
 BQ682356.1 GI:21795035
KEYWORDS
 EST.
SOURCE
 human.
ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
 1 (bases 1 to 1046)
AUTHORS
 NIH-MGC http://mgc.nci.nih.gov/.
TITLE
 National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
 Unpublished (1999)
COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: DCTD/BTP

CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LCM2420 row: m column: 17
 High quality sequence stop: 231.
FEATURES
 Location/Qualifiers
 1..1046
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:6260920"
 /clone_lib="NIH_MGC_112"
 /tissue_type="melanotic melanoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."
 BASE COUNT 46 a 401 c 203 g 39 t 357 others
 ORIGIN
 Query Match 4.9%; Score 49.4; DB 14; Length 1046;
 Best Local Similarity 23.0%; Pred. No. 0.024;
 Matches 161; Conservative 0; Mismatches 538; Indels 0; Gaps 0;
 QY 97 CCCCTGCTGCTCTGTCCTACCCAGGCCCCCAACCCCCCCCCCAGCAGCTGTTC 156
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 851 CGCGCGCAGCG 792
 QY 157 AGGCTCTCAGCTGCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 216
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 791 CG 732
 QY 217 CAGGCCAGGCGAGTGGTGTGGAGTGAAGAGCAGACGCGGTGGGGAGGTTCAGAGAGA 276
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 731 GGG 672
 QY 277 ATCTGCTGGGTGGGATGGTGTGGGCATCACTGTCCCATCTGCTGACAGCTGCT 336
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 671 GGGGGGGGNGGG 612
 QY 337 GGCAGGAAGGAGTGGGGGGCCATAGCAGTGTGTGTCACCGAGGCTGGCCTGGAAGTG 396
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 611 GGG 552
 QY 397 GTGCCAGCACTACTAAGAGCCAGGAAAGCCCTGCCAAGTGTGTGGCCCTAGTTC 456
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 551 NNN 492
 QY 457 TCATCAGCGCCCTAGCAGCGCCCTGTGTCTCAGGTAAGGGGAGGCTGTAGCACA 516
 Db 491 NNN 432
 QY 517 TAGTCAGCCCTGGTGTCCCATGCTTCTCTCTCTGTGCCCAATTTTAGGGCCATGT 576
 Db 431 NNN 372
 QY 577 ATTTGGGGCTATGTGACTCATGCTGTAAGGTGCTTGGGCCAGAGCTGTGGGCACCTTT 636
 Db 371 NNN 312
 QY 637 AAATGCCACCCAGTCTCATGTGCCGAGTTTGGGGTAGGCTAGGTAGGATTGGGAATA 696
 Db 311 NNN 252
 QY 697 TGGGAGGAGCGAGGATCTGTCTACCTAGGAGGCACTCTCTCTCTCTCTCTCTCT 756
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db      251 NNNNNNNNNNNNNNNNNNNNNNNCCACTCCCTCACCATCGTCTTGCGGTGCGRGA 192

QY      757 CAAGAGAACTGAACGTTGGTAGGGGCTCAGGACGATG 795
          |   |   |   |   |   |   |   |   |   |
Db      191 GCTAGGCTCAGGCTGGAGGCTCGGGTCCTGACGTCAGTG 153


RESULT 5
BM479595/c
LOCUS    BM479595
DEFINITION     1398 bp mRNA linear EST 05-FEB-2002
              5'.. mRNA sequence.
ACCESSION     AGENCOURT_6418829 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5502428
VERSION       BM479595
KEYWORDS      BM479595.1 GI:18528637
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1' (bases 1 to 1398)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-re@mail.nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
Plate: LLAM12140 row: 1 column: 21
High quality sequence stop: 225.
FEATURES             Location/Qualifiers
     source            1..1398
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5502428"
                     /clone_lib="NIH_MGC_67"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 Kb. Library constructed by Life
                     Technologies."
BASE COUNT           191 a 669 c 124 g 410 t 4 others
ORIGIN
Query Match               4.8%; Score 48.4; DB 13; Length 1398;
Best Local Similarity 52.5%; Pred. No. 0.048;
Matches 106; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY      184 GTCTGCCCTGGGAGAAATAGGTGGGAGAAACAGCCAGGCAGTGTGTTGGAGTG 243
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      512 GTGGGGGTGGAAAAAGGAGGAGGGGGGTAAAAGAAGGGGGGTAGAAAATGGAGGG 453
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      244 AAGCAGACAGCGGTGGGAGGTCAGAGAGATCTCTGGGCTGGGGATGTTGGGCG 303
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      452 GAGGGGGAGGGGAAGGGGGGAAGGGGGGAAAAAAGGGGGGAGAGAGGTTGGGGGG 393
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      304 ATCAACTGTCCCATTGCTGCAGGCTGTCTTTGGGCGAGGGAAGGGGATGGGGCCCATAG 363
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      392 ATGATGATAGTGTTCAGGCTGGGGAACCTGGGGGGGAATAGGGGGGGGGGAAGGG 333
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      364 CAGTGTGTTGTCAGCCAGGCTGG 385
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      332 GCGGGGGACAGGGCAGGAGCG 311
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6		
CNS04IN8/c		
LOCUS :	CNS04IN8	DNA
	817 bp	linear
		GSS 21-MAY-2000

DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 112H08 of library G from Tetraodon nigroviridis, genomic survey sequence.									
ACCESSION	AL292445									
VERSION	AL292445.1 GI:8031025									
KEYWORDS	GSS; genome survey sequence.									
SOURCE	Tetraodon nigroviridis									
ORGANISM	Tetraodon nigroviridis									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.									
REFERENCE	1 (bases 1 to 817)									
AUTHORS	Roest-Croliius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.									
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence									
JOURNAL	Unpublished									
REFERENCE	2 (bases 1 to 817)									
AUTHORS	Roest-Croliius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.									
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis									
JOURNAL	Unpublished									
REFERENCE	3 (bases 1 to 817)									
AUTHORS	Genoscope.									
TITLE	Direct Submission									
JOURNAL	Submitted (12-APR-2000)									
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .									
FEATURES	<p>source</p> <p>1. 817</p> <p>Location/Qualifiers</p> <p>/organism="Tetraodon nigroviridis"</p> <p>/db_xref="taxon:99883"</p> <p>/clone="112H08"</p> <p>/clone_1lb="G"</p> <p>/note="Genoscope sequence ID : C08G112D04LPL-end : T7"</p>									
BASE COUNT	157 a 335 c 107 g 135 t 83 others									
ORIGIN										
	<p>Query Match 4.88; Score 47.8; DB 17; Length 817;</p> <p>Best Local Similarity 44.18; Pred. No. 0.059;</p> <p>Matches 123; Conservative 16; Mismatches 140; Indels 0; Gaps 0;</p>									
Qy	94 CACCCCTCTGCTTCCTGGCCTCACCCAGGCCCCCAACCCCGCCGCCCCAGCAGCTGTT 153									
Db	707 CTCCCCCCCACCTCTCATTTACCTACCTAATCTSTTCATATTTTAKSSCTTTSAKSDST 648									
Qy	154 CTCAGGCGCTCAGCCCTGTCGATTCCTGCTGCGCCCTGGGGAAGATGAGTGGGAGAA 213									
Db	647 WTATTCGCGTAMTCYCTCKATCATCTATTATSGGGGGGGGGGGGGGGGGGGGGGA 588									
Qy	214 AACCCAGCCAGGCGAGTTGTTGGAGTGAAGACAGACGCGGTGGGAGGTCAGGAG 273									
Db	587 GTAGSGGGGGGAGGGGAGGGGGGGGGGAGGAGGAGGGGGGGGGGGAAGAGAA 528									
Qy	274 AGAATCTGCTGGGCTGGGGATGTTGTGGGCATCAACTTGTCCCATTTGTCAGGCTGGTCT 333									
Db	527 AGGGGGGGTGGGGGGGGTGGGGGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAG 468									
Qy	334 TGGGCGAGGAAGGGGATGGGGGCCATACGAGTGTGG 372									
Db	467 GGGGCTGGGGGGGGGGGGGGGGGGGGAAGGGGGGAGG 429									

RESULT 7	LOCUS	910 bp	DNA	linear	GSS 24-SEP-2001
BH162948	BH162948				
DEFINITION	ENTR49TR	Entamoeba histolytica	sheared	DNA	Entamoeba histolytica


```

Db      150 GGGGTGKRRARGGGGRKKRRTKG 122          |||.....:::||

RESULT 13
LOCUS   CNS006UO                                884 bp    DNA        linear     GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACKI4N21 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL065923
VERSION   AL065923.1 GI:4944891
KEYWORDS  GSS.
SOURCE    Drosophila melanogaster.
           Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epitheroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 884)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr ; Web : www.genoscope.cns.fr )
COMMENT   Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library.html prepared by Kazutoyo Osogawa and Aaron Mammoler at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
     source            1..884
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACRI4N21"
                     /clone_lib="RPCI-98"
                     /notes=end : T7"

BASE COUNT       230 a      62 c      139 g      124 t      329 others

ORIGIN
Query Match               4.6%; Score 46; DB 17; Length 884;
Best Local Similarity    19.5%; Pred. No. 0.18;
Matches                  44; Conservative 93; Mismatches 89; Indels 0; Gaps 0;

QY  174 TGATTTCCTTGTCTGGCCTGGGAGAAATGAGTGCGGAGAACCAGCCAGGCAGTTGG 233
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  623 TGTKTGTGTTKKKGKGGGGGGGGKKGKTGKTGKGKGGGGGGGGKKGKTGKTGKGGK 682
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  234 TTGTGGAGTGAAGACAGACGCCGGTGGGGAGGTCAAGAGAGAATCTGTGGCTGGGGA 293
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

DB  683 KKKKKKGTGKGNKGTTKKSXTFKKKGGTGKTGTGVTGTTTTKKKGKGGKGGKKNKGGKR 742
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  294 TGGTGTGGCATCAACTGTCCCATTGCTGCAGGCTGGTCTTGGGGCAGGAAGGGATGG 353
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

DB  743 KGKKGKGGKGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 802
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  354 GGCGCCATACAGTGTGCTCACCCAGGCTGGCCTGGGAAGTGCTG 399
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

DB  803 KKRGKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 848
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
AL543262/C
LOCUS    AL543262                951 bp    mRNA        linear     EST 16-FEB-2001
```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:00:38 ; Search time 2122.53 Seconds
(without alignments)
13725.086 Million cell updates/sec

Title: US-09-691-220-3_COPY_10000_11000
Perfect score: 1001
Sequence: 1 cccaggagactgcagctgg.....aaggctgcctgcctggacat 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:**

- 1: gb_ba:**
- 2: gb_htg:**
- 3: gb_in:**
- 4: gb_om:**
- 5: gb_ov:**
- 6: gb_pat:**
- 7: gb_ph:**
- 8: gb_pl:**
- 9: gb_pr:**
- 10: gb_ro:**
- 11: gb_sts:**
- 12: gb_sy:**
- 13: gb_un:**
- 14: gb_vi:**
- 15: em_ba:**
- 16: em_fun:**
- 17: em_hum:**
- 18: em_in:**
- 19: em_mu:**
- 20: em_om:**
- 21: em_or:**
- 22: em_ov:**
- 23: em_pat:**
- 24: em_ph:**
- 25: em_pl:**
- 26: em_ro:**
- 27: em_sts:**
- 28: em_un:**
- 29: em_vi:**
- 30: em_htg_hum:**
- 31: em_htg_inv:**
- 32: em_htg_other:**
- 33: em_htg_mus:**
- 34: em_htg_pln:**
- 35: em_htg_rod:**
- 36: em_htg_man:**
- 37: em_htg_vrt:**
- 38: em_sy:**
- 39: em_htgo_hum:**
- 40: em_htgo_mus:**
- 41: em_htgo_other:**

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	1001	100.0	138999	9	AC090426	AC090426 Homo sapi
2	1001	100.0	158766	2	AC015851	AC015851 Homo sapi
3	1001	100.0	166368	2	AC018629	AC018629 Homo sapi
C 4	1001	100.0	173441	2	AC080112	AC080112 Homo sapi
C 5	1001	100.0	188574	2	AC131063	AC131063 Homo sapi
6	1001	100.0	190309	2	AC126392	AC126392 Homo sapi
7	997.8	99.7	1233	9	HSRARA1A5	AF088892 Homo sapi
8	255.6	25.5	231248	2	AL591067	AF088892 Homo sapi
C 9	204.4	20.4	162393	2	AC111746	AC111746 Rattus no
10	174	17.4	704	6	AR052138	AR052138 Sequence
11	174	17.4	1284	9	S50916	S50916 PML-RAR fus
12	174	17.4	1572	12	AF242868	AF242868 Synthetic
13	174	17.4	1920	9	HSRAR	X06538 Human mRNA
14	174	17.4	1944	9	HSU41743	U41743 Human nucle
15	174	17.4	2008	9	AK098172	AK098172 Homo sapi
16	174	17.4	2073	9	HSU41742	U41742 Human nucle
17	174	17.4	2085	12	AF242867	AF242867 Synthetic
18	174	17.4	2441	9	BC008727	BC008727 Homo sapi
19	174	17.4	2658	6	AR052137	AR052137 Sequence
20	174	17.4	2907	6	AX333042	AX333042 Sequence
21	174	17.4	2907	9	HSRRA	X06614 Human mRNA
22	174	17.4	2928	6	AR061122	AR061122 Sequence
23	174	17.4	2940	6	AR052136	AR052136 Sequence
24	174	17.4	2940	6	I09348	I09348 Sequence 1
25	174	17.4	3036	6	AR061121	AR061121 Sequence
26	174	17.4	3036	6	I96215	I96215 Sequence 52
27	174	17.4	3036	9	HUMPLRAR	M73779 Human PML-R
28	172.4	17.2	2908	6	I08117	I08117 Sequence 1
29	167.6	16.7	1474	10	MSRARA	Y10094 M.spretus m
30	167.6	16.7	1759	10	MURARAP	M60909 Mouse retin
31	167.6	16.7	1926	10	NMRARA	X57528 M.musculus
32	167.6	16.7	2061	10	S56656	S56656 retinolic ac
33	166	16.6	908	10	RNAJ2940	AJ002940 Rattus no
34	166	16.6	924	10	RNAJ2941	AJ002941 Rattus no
35	166	16.6	2130	10	RNU15211	U15211 Rattus norv
36	162.8	16.3	2748	10	AY046943	AY046943 Mesocric
C 37	161.4	16.1	2300	9	HSR05408	AL834159 Homo sapi
38	156	15.6	301	6	AX098489	AX098489 Sequence
39	151.8	15.2	392	4	SSU82260	U82260 Sus scrofa
40	143.2	14.3	1494	5	GGRARA21	X78335 G.gallus re
41	143.2	14.3	1552	5	GGRARA1	X78372 G.gallus RA
42	138.8	13.9	860	5	AB073783	AB073783 Paralicht
43	137.8	13.8	1377	5	NVRARAG	X17585 Newt mRNA f
44	137.8	13.8	1697	5	NVRARA2M	Z14254 N.viridesc
C 45	132.2	13.2	110000	2	AC021072_2	Continuation (3 of

ALIGNMENTS

RESULT 1
AC090426/c
LOCUS
DEFINITION Homo sapiens chromosome 17 clone 205m17 map 17q21.1, complete
sequence.
ACCESSION AC090426
VERSION AC090426.1 GI:12965315
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 138999)
AUTHORS Gu,B., Xiong,H., Zhou,Y., Chen,B., Lu,L., Zhong,M., Yin,H.,
Huang,W., Ren,S., Chen,S.F., Chen,Z. and Fu,G.

TITLE	Variant-type PML-RAR(alpha) fusion transcript in acute promyelocytic leukemia: use of a cryptic coding sequence from intron 2 of the RAR(alpha) gene and identification of a new clinical subtype resistant to retinoic acid therapy
JOURNAL MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 99 (11), 7640-7645 (2002)
PUBMED	22028997
REFERENCE	12032336
AUTHORS	2 (bases 1 to 138999) Gu,B., Xiong,H., Zhou,Y., Chen,B., Lu,L., Zhong,M., Yin,H., Huang,W., Ren,S., Chen,S.F., Chen,Z. and Fu,G.
TITLE	Direct Submission
JOURNAL	Submitted (20-FEB-2001) Chinese National Human Genome Center at Shanghai
COMMENT	Shanghai, Shanghai, Shanghai 201203, P.R.China -----Genome Center----- Center:Chinese National Human Genome Center at Shanghai Center code:CHGC Website: http://www.chgc.sh.cn Contact: fufang@chgc.sh.cn.
FEATURES	Location/Qualifiers 1..138999 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="17" /map="17q21.1" /clone="205m17"
repeat_region	387..483 /rpt_family="SINE/MIR" /rpt_unit="MIR" 725..1032 /rpt_family="SINE/Alu" /rpt_unit="AluJb" 1033..1198 /rpt_family="SINE/Alu" 1217..1514 /rpt_family="SINE/Alu" /rpt_unit="AluSp complement(1534..1754) /rpt_family="SINE/MIR" /rpt_unit="MIR" 2291..2366 /rpt_family="SINE/MIR" /rpt_unit="MIR complement(2607..2665) /rpt_family="SINE/MIR" /rpt_unit="MIR 3631..3905 /rpt_family="LINE/L1" /rpt_unit="L1ME2 3977..4029 /rpt_family="LINE/L1" /rpt_unit="L1ME2 4030..4213 /rpt_family="LTR/MaLR" /rpt_unit="MSTD 4214..4509 /rpt_family="SINE/Alu" /rpt_unit="AluJb 4514..4811 /rpt_family="SINE/Alu" /rpt_unit="AluSx 4812..5018 /rpt_family="LTR/MaLR" /rpt_unit="MSTD 5019..5102 /rpt_family="LINE/L1" /rpt_unit="L1ME2 5236..5443 /rpt_family="SINE/MIR" /rpt_unit="MIR 5467..5699 /rpt_family="SINE/MIR" /rpt_unit="MIR 6547..6708 /rpt_family="SINE/MIR"
repeat_region	6827..6917 /rpt_family="Low_complexity" /rpt_unit="GA-rich 7733..7779 /rpt_family="SINE/MIR" /rpt_unit="MIR 8268..8411 /rpt_family="Simple_repeat" /rpt_unit="(TGGG)n 9224..9532 /rpt_family="SINE/Alu" /rpt_unit="AluY 9546..9639 /rpt_family="Low_complexity" /rpt_unit="CT-rich 10025..10426 /rpt_family="LTR/Retroviral" /rpt_unit="MLT2A 10433..10471 /rpt_family="Simple_repeat" /rpt_unit="(TA)n 10472..10527 /rpt_family="LTR/Retroviral" /rpt_unit="MLT2A complement(10675..10962) /rpt_family="SINE/Alu" /rpt_unit="AluY 11013..11321 /rpt_family="SINE/Alu" /rpt_unit="AluY complement(11840..12192) /rpt_family="LTR/MaLR" /rpt_unit="MSTD 12729..12835 /rpt_family="SINE/Alu" /rpt_unit="FLAM_A complement(12841..12891) /rpt_family="LINE/L2" /rpt_unit="L2 complement(14228..14548) /rpt_family="SINE/Alu" /rpt_unit="AluY complement(14618..14701) /rpt_family="SINE/MIR" /rpt_unit="MIR 14953..15255 /rpt_family="SINE/Alu" /rpt_unit="AluJo 15311..15473 /rpt_family="Simple_repeat" /rpt_unit="(TA)n 15669..15981 /rpt_family="SINE/Alu" /rpt_unit="AluSx 16316..16353 /rpt_family="Simple_repeat" /rpt_unit="(TAAA)n complement(16863..16910) /rpt_family="SINE/MIR" /rpt_unit="MIR 17026..17324 /rpt_family="SINE/Alu" /rpt_unit="AluSg 17337..17618 /rpt_family="SINE/Alu" /rpt_unit="AluSx 17619..18024 /rpt_family="LTR" /rpt_unit="LTR54 complement(18026..18086) /rpt_family="SINE/MIR" /rpt_unit="MIR

D	b	155184	AGATCTCTGGGCTGTCCTTCCCTTTCCCCTCTCTCTTCCTCCTCTCTGTGCCTCTTCCTCAAGG	155125
Q	y	721	AGCTCCCGAGGAAGTGAAAGCGTGGTAGAGGCGAGGCGCTGTGGGGCTGGAGCCAGCGCTGA	780
D	b	155124	AGCTCCCGAGGAAGTGAAAGCGTGGTAGAGGCGAGGCGCTGTGGGGCTGGAGCCAGCGCTGA	155065
Q	y	781	GAGGGGTGCCATGGAGGAAGACGCCCTCACCTCTCCCTCTCCCCCAGAACACAGCTCA	840
D	b	155064	GAGGGGTGCCATGGAGGAAGACGCCCTCACCTCTCCCTCTCCCCCAGAACACAGCTCA	155005
Q	y	841	GAACAAGCTGTCTCTCTGGACATTGACCTCTGGGACAACTTCAGTGAACCTCCACCACAG	900
D	b	155004	GAACAAGCTGTCTCTCTGGACATTGACCTCTGGGACAACTTCAGTGAACCTCCACCACAG	154945
Q	y	901	TGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCACCAACCTCCACCATC	960
D	b	154944	TGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCACCAACCTCCACCATC	154885
Q	y	961	GCGACACAGATCACCCCTCTCAAGGCTGCCTGGACAT	1001
D	b	154884	GCGACACAGATCACCCCTCTCAAGGCTGCCTGGACAT	154844
R	E	S	A	C
RESULT	6	AC126392	190309 bp	DNA linear HTG 06-AUG-2001
LOCUS	AC126392	Homo sapiens chromosome 17 clone RP11-1029F16 map 17, WORKING DRAFT	SEQUENCE, 11 ordered pieces.	
DEFINITION	AC126392	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.	human.	
ACCESSION	VERSION	AC126392.2	GI:22123095	
KEYWORDS	FEATURES	key words:		
SOURCE	ORGANISM	Homo sapiens		
REFERENCE	AUTHORS	Bukacinski, J., et al.		
TITLE	JOURNAL	Homo sapiens chromosome 17, clone RP11-1029F16		
REFERENCE	AUTHORS	Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgaiter, B., Camarata, J., Chang, J., Chazaro, B., Choelpel, Y., Collymore, A., Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Milenga, V., Murphy, T., Navlor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Sever, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemlek, L., Zimmer, A. and Zody, M.		
Direct Submission				
TITLE	JOURNAL	Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	AUTHORS	Research, 320 Charles Street, Cambridge, MA 02141, USA		
(bases 1 to 190309)				
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgaiter, B., Camarata, J., Chang, J., Chazaro, B., Choelpel, Y., Collymore, A., Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Milenga, V., Murphy, T., Navlor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Sever, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemlek, L., Zimmer, A. and Zody, M.				

Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:21699357.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L27553

Center clone name: 1029_F_16

Summary Statistics

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 187521 bases at least Q40

Consensus quality: 188193 bases at least Q30

Consensus quality: 188537 bases at least Q20

Insert size: 179000; agarose-fp

Insert size: 189309; sum-of-contigs

Quality coverage: 15.5 in Q20 bases; agarose-fp

Quality coverage: 14.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently

* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 30513: contig of 30513 bp in length

* 30514 30613: gap of 100 bp

* 30614 31764: contig of 1151 bp in length

* 31765 31864: gap of 100 bp

* 31865 33693: contig of 1829 bp in length

* 33694 33793: gap of 100 bp

* 33794 35405: contig of 1612 bp in length

* 35406 35505: gap of 100 bp

* 35506 37691: contig of 2186 bp in length

* 37692 37791: gap of 100 bp

* 37792 51370: contig of 13579 bp in length

* 51371 51470: gap of 100 bp

* 51471 66014: contig of 14544 bp in length

* 66015 66114: gap of 100 bp

* 66115 108183: contig of 42069 bp in length

* 108184 108283: gap of 100 bp

* 108284 130510: contig of 22227 bp in length

* 130511 130610: gap of 100 bp

* 130611 166740: contig of 36130 bp in length

* 166741 166840: gap of 100 bp

* 166841 190309: contig of 23469 bp in length.

Location/Qualifiers

1. 190309

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosomes="17"

/map="17"

/clone="RP11-1029F16"

/clone_lib="RPC1-11 Human Male BAC"

FEATURES
source

misc_feature 1. 30513
clone_end:SP6
vector_side:left
30614. 31764
/note="assembly_fragment"
31865. 33693
/note="assembly_fragment"
33794. 35405
/note="assembly_fragment"
35506. 37691
/note="assembly_fragment"
37792. 51370
/note="assembly_fragment"
51471. 66014
/note="assembly_fragment"
66115. 108183
/note="assembly_fragment"
108284. 130510
/note="assembly_fragment"
130611. 166740
/note="assembly_fragment"
166841. 190309
/note="assembly_fragment"
clone_end:T7
vector_side:right"

BASE COUNT 48207 a 48484 c 47543 g 45018 t 1057 others
ORIGIN

Query Match 100.0%; Score 1001; DB 2; Length 190309;
Best Local Similarity 100.0%; Pred. No. 5.2e-206;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCCAGGAGACTGCAGCTGGGAGGCTGGGTGAGTGGAGGCGGAGAGAGACCTTCCTGG 60
|||||
Db 46671 CCCAGGAGACTGCAGCTGGGAGGCTGGGTGAGTGGAGGCGGAGAGAGACCTTCCTGG 46730
Qy 61 GGAAAGAGAGGAGAGAGACCTTAGAGGGCACCCGCTGCGCTGGAGTGTGAGCTGGAGTAGA 120
|||||
Db 46731 GGAAAGAGAGGAGAGAGACCTTAGAGGGCACCCGCTGCGCTGGAGTGTGAGCTGGAGTAGA 46790
Qy 121 CGCGTGGGGATAGCATCGCGCTGCCTATAGGGTGGGGTGGGGTGGTGTGAGGGCCAC 180
|||||
Db 46791 CGCGTGGGGATAGCATCGCGCTGCCTATAGGGTGGGGTGGGGTGGTGTGAGGGCCAC 46850
Qy 181 AGCTGTCTCATGGGCTTCTGGGCGAGACTTGATGTGTGGTGGGTGGGCATGGAGG 240
|||||
Db 46851 AGCTGTCTCATGGGCTTCTGGGCGAGACTTGATGTGTGGTGGGTGGGCATGGAGG 46910
Qy 241 GCTGGAGTGCCTGGCAATGCCCTTGCCTCCCGTGAACGCGTGTCTGTGCGCGTGTCTTAC 300
|||||
Db 46911 GCTGGAGTGCCTGGCAATGCCCTTGCCTCCCGTGAACGCGTGTCTGTGCGCGTGTCTTAC 46970
Qy 301 AAGCCTGGGTGACCTCTCAGCAGCTGCGAGCTCTCTGTGAGGCTGGGGTGGAGCAGGC 360
|||||
Db 46971 AAGCCTGGGTGACCTCTCAGCAGCTGCGAGCTCTCTGTGAGGCTGGGGTGGAGCAGGC 47030
Qy 361 CCTGAGCAGCTGAGCTGCCCTCTTAAACCCCTCTGCTCCACAGCTGTGAGAAACA 420
|||||
Db 47031 CCTGAGCAGCTGAGCTGCCCTCTTAAACCCCTCTGCTCCACAGCTGTGAGAAACA 47090
Qy 421 CCGAAACAAGAAGAAGAGAGGTGCCCAAGCCGAGTGTCTGTAGAGACTACACCTGTAC 480
|||||
Db 47091 CCGAAACAAGAAGAAGAGAGGTGCCCAAGCCGAGTGTCTGTAGAGACTACACCTGTAC 47150
Qy 481 GCCGGAGTGGGGAGCTCATTGAGAGGTGCGCAAAAGCCGAGGAAACCTTCCTCCCTGC 540
|||||
Db 47151 GCCGGAGTGGGGAGCTCATTGAGAGGTGCGCAAAAGCCGAGGAAACCTTCCTCCCTGC 47210
Qy 541 CCTCTGCCAGCTGGGCAAAATACACTAGGATAGGCTTTCCCGCCGCTGCAGGGTGGAT 600
|||||
Db 47211 CCTCTGCCAGCTGGGCAAAATACACTAGGATAGGCTTTCCCGCCGCTGCAGGGTGGAT 47270

[illegible]

Query Match 99.7%; Score 997.8; DB 9; Length 1233;
Best Local Similarity 99.8%; Pred. No. 4.7e-205;
Matches 999; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

[illegible]

RESULT 8
AL591067
LOCUS AL591067 231248 bp DNA 1 linear HTG 09-AUG-2002
DEFINITION Mus musculus chromosome 11 clone RP23-333D2, *** SEQUENCING IN

Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, J., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
 Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, K.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 162393)
 Worley, K.C.
 Direct Submission
 Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 162393)
 Worley, K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:18701588.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GOAF
 Center clone name: CH230-15011
 ----- Summary Statistics
 : Sequencing vector: Plasmid;
 : Chemistry: Dye-terminator Big Dye: 100% of reads
 : Assembly program: Phrap; version 0.990329
 : Consensus quality: 96609 bases at least Q40
 : Consensus quality: 101360 bases at least Q30
 : Consensus quality: 105362 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * : (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 72 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1
 * 1242: contig of 1242 bp in length
 * 1243: gap of unknown length
 * 1343: contig of 1085 bp in length
 * 2428: gap of unknown length
 * 2527: gap of unknown length
 * 2528: contig of 1108 bp in length
 * 3635: gap of unknown length
 * 3735: gap of unknown length
 * 4821: contig of 1086 bp in length
 * 4822: gap of unknown length
 * 4921: gap of unknown length
 * 4922: contig of 1719 bp in length
 * 6641: gap of unknown length
 * 6741: contig of 1476 bp in length
 * 8216: gap of unknown length
 * 8316: gap of unknown length
 * 9317: contig of 1030 bp in length
 * 9346: gap of unknown length
 * 9447: contig of 1245 bp in length
 *
 * 10692
 * 10792
 * 11851: contig of 1060 bp in length
 * 11951: gap of unknown length
 * 11952: contig of 1558 bp in length
 * 13510: gap of unknown length
 * 13609: gap of unknown length
 * 14869: contig of 1260 bp in length
 * 14969: gap of unknown length
 * 14970: contig of 1306 bp in length
 * 16276: gap of unknown length
 * 16375: gap of unknown length
 * 17440: contig of 1065 bp in length
 * 17441: gap of unknown length
 * 17541: gap of unknown length
 * 18957: gap of unknown length
 * 20061: contig of 1005 bp in length
 * 20062: gap of unknown length
 * 20162: gap of unknown length
 * 21232: contig of 1071 bp in length
 * 21332: gap of unknown length
 * 21333: gap of unknown length
 * 22378: contig of 1046 bp in length
 * 22379: gap of unknown length
 * 23629: contig of 1151 bp in length
 * 23630: gap of unknown length
 * 23729: gap of unknown length
 * 23730: contig of 1264 bp in length
 * 24994: gap of unknown length
 * 25094: gap of unknown length
 * 26649: gap of unknown length
 * 26749: gap of unknown length
 * 28384: gap of unknown length
 * 28484: gap of unknown length
 * 29659: gap of unknown length
 * 29759: gap of unknown length
 * 31549: gap of unknown length
 * 31649: gap of unknown length
 * 32916: gap of unknown length
 * 33015: gap of unknown length
 * 34132: contig of 1117 bp in length
 * 34133: gap of unknown length
 * 34233: gap of unknown length
 * 35377: gap of unknown length
 * 35477: gap of unknown length
 * 36835: contig of 1359 bp in length
 * 36836: gap of unknown length
 * 36936: gap of unknown length
 * 38119: contig of 1184 bp in length
 * 38219: gap of unknown length
 * 38220: gap of unknown length
 * 40048: contig of 1829 bp in length
 * 40049: gap of unknown length
 * 40149: gap of unknown length
 * 41648: contig of 1500 bp in length
 * 41649: gap of unknown length
 * 41748: gap of unknown length
 * 44023: contig of 2275 bp in length
 * 44024: gap of unknown length
 * 44123: gap of unknown length
 * 44124: gap of unknown length
 * 45297: contig of 1174 bp in length
 * 45298: gap of unknown length
 * 45398: gap of unknown length
 * 45399: contig of 1136 bp in length
 * 46534: gap of unknown length
 * 46634: gap of unknown length
 * 48452: contig of 1818 bp in length
 * 48551: gap of unknown length
 * 48552: gap of unknown length
 * 49889: contig of 1338 bp in length
 * 49890: gap of unknown length
 * 49990: gap of unknown length
 * 52066: contig of 2077 bp in length
 * 52067: gap of unknown length
 * 52167: gap of unknown length
 * 53724: contig of 1558 bp in length
 * 53725: gap of unknown length
 * 53825: gap of unknown length
 * 53825: contig of 1526 bp in length
 * 55351: gap of unknown length
 * 55450: gap of unknown length
 * 55451: gap of unknown length
 * 57835: contig of 2385 bp in length
 * 57836: gap of unknown length
 * 57935: gap of unknown length
 * 59470: contig of 1535 bp in length
 * 59471: gap of unknown length
 * 59570: gap of unknown length
 * 61049: contig of 1479 bp in length
 * 61050: gap of unknown length
 * 61150: gap of unknown length
 * 62619: contig of 1470 bp in length
 * 62620: gap of unknown length
 * 62719: gap of unknown length
 * 65125: contig of 2406 bp in length
 * 65126: gap of unknown length
 * 65225: gap of unknown length
 * 65226: gap of unknown length
 * 66427: contig of 1202 bp in length
 * 66428: gap of unknown length

[illegible]

Db 96860 T 96860

RESULT 10
AR052138 LOCUS AR052138 704 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 5 from patent US 5830760;
ACCESSION AR052138
VERSION AR052138.1 GI:5975502
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
UNCLASSIFIED.
1 (bases 1 to 704)
Tsai,S. and Collins,S.J.
TITLE Creating novel hematopoietic cell lines by expressing altered
retinoic acid receptors
JOURNAL Patent: US 5830760-A 5 03-NOV-1998;
FEATURES
source location/Qualifiers
1. .704
/organism="unknown"
BASE COUNT 150 a 232 c 202 g 120 t
ORIGIN

Query Match 17.4%; Score 174; DB 6; Length 704;
Best Local Similarity 100.0%; Pred. No. 2.7e-27;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACAAAGCTGTCTCTCTGGACATTGACCTCTGGGACAAGTTCAGTGA 887
|||||
DB 41 GAACAACAGCTCAGAACAAAGCTGTCTCTCTGGACATTGACCTCTGGGACAAGTTCAGTGA 100
|||||
QY 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 947
|||||
DB 101 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 160
|||||
QY 948 CACCCTCACCATCGCCGACCAGATCACCCCTCTCAAGGCTGCCTGCCTGGACAT 1001
|||||
DB 161 CACCCTCACCATCGCCGACCAGATCACCCCTCTCAAGGCTGCCTGCCTGGACAT 214
|||||

RESULT 11
S50916
LOCUS S50916 1284 bp mRNA linear PRI 29-JUN-2000
DEFINITION PML-RAR fusion gene [fusion transcript] [human, mRNA Partial, 1284 nt].
ACCESSION S50916
VERSION S50916.1 GI:234248
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1284)
de The,H., Lavau,C., Marchio,A., Chomienne,C., Degos,L. and Dejean,A.
TITLE The PML-RAR alpha fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR
JOURNAL Cell 66 (4), 675-684 (1991)
MEDLINE 91347369
PUBMED 1652369
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gbbsq 50916] from the original journal article. This sequence comes from fig2b.
FEATURES
source Location/Qualifiers
1. .1284
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1. .1284
/gene="PML-RAR fusion gene"
<1. .1251

FTTLTIADQITLLKAAACLDLILRICITRYTPEQDTWTFSDGLTLNRTOMHNAGFGPLT
DLVAFANQLLPLEMDDAETGELLACICGDRQDLEQDPRVMDLQEPLELALKAYVIR
KRPSRPHMPKMLKITDLSISAKGAERVITLTKMEIPGSMPLIQEMLNSENGLDTR
LSQPGGGGDRGGGLAPPSPSCPSLSPSSNRSSPATHSP"

misc_feature

old_sequence

misc_feature

old_sequence

old_sequence

BASE COUNT
ORIGIN

Query Match 17.4%; Score 174; DB 9; Length 1920;
Best Local Similarity 100.0%; Pred. No. 2.4e-27;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAGTTCACTGA 887

Db 856 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAGTTCACTGA 915

QY 888 ACTCTCACCAGTGCATTAAGACTGTGAGTTCGCCAAGCAGCTGCCGGCTTCAC 947

Db 916 ACTCTCACCAGTGCATTAAGACTGTGAGTTCGCCAAGCAGCTGCCGGCTTCAC 975

QY 948 CACCTTCACCATCGCCGACAGATCACCTCTCAAGGCTGCCTGCCTGGACAT 1001

Db 976 CACCTTCACCATCGCCGACAGATCACCTCTCAAGGCTGCCTGCCTGGACAT 1029

RESULT 14

HSU41743

LOCUS

DEFINITION Human nucleophosmin-retinoic acid receptor alpha fusion protein
NPM-RAR short form mRNA, complete cds.

ACCESSION U41743

VERSION U41743.1

KEYWORDS GI:1314309

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE 1

AUTHORS Redner, R.L., Rush, E.A., Faas, S., Rudert, W.A. and Corey, S.J.

TITLE The t(5;17) variant of acute promyelocytic leukemia expresses a

JOURNAL nucleophosmin-retinoic acid receptor fusion

MEDLINE Blood 87 (3), 882-886 (1996)

PUBMED 96151966

REFERENCE 2

AUTHORS Redner, R.L.

TITLE Direct Submission

JOURNAL Submitted (30-NOV-1995) Robert L. Redner, Medicine, University of

FEATURES

Source

Location/Qualifiers

1. 1944

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="acute promyelocytic leukemia cells with

t(5;17)(q32;q11) translocation"

/tissue_type="bone marrow"

84..1646

/codon_start=1

/product="nucleophosmin-retinoic acid receptor alpha

fusion protein NPM-RAR short form"

CDS

/protein_id="AAB00113.1"
/db_xref="GI:1314310"
/translation="MEDSDMDMSPLRPQNYLFGCELKADKYHFVDMDENERQSLSL
KRVSLGAKADLHIVAEAMNVEGSPIRKVTATLTKMSQVPTVSLGFEITPPVVLRL
KCGSGPVHISGQHLVAIETQSSSEIIVSPSPPLPRIYKPCFVCQDKSSSYHYGV
SACECKGFFRRSIQKNMVTCHURKNCLINRVNRQCTRLQKCFEVMGSKESVRN
DRNKKKEVPKPESESYYLTTPVEGELIEBKVRKHOETEPALCOLQKYTTNNSEQVR
SLIDILWDKFSLSLTKCIITVEFAQLQFGLTLDVAFANQLLPLEMDDAETGLLSAICL
YTPEDQTMFTSDGLTLNRTOMHNAGFGPLTDLVAFANQLLPLEMDDAETGLLSAICL
ICGDRQDLEQDPRVMDLQEPLELALKAYVIRKRPSRPHMPKMLKITDLSISAKGA
ERVITLTKMEIPGSMPLIQEMLNSENGLDTRLSQPGGGGDRGGGLAPPSPSCPSLSP
SSNRSSPATHSP"

BASE COUNT 427 a 598 c 557 g 362 t

ORIGIN

Query Match 17.4%; Score 174; DB 9; Length 1944;

Best Local Similarity 100.0%; Pred. No. 2.4e-27;

Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAGTTCACTGA 887

Db 887 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAGTTCACTGA 946

QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGAGTTCGCCAAGCAGCTGCCGGCTTCAC 947

Db 947 ACTCTCCACCAAGTGCATTAAGACTGTGAGTTCGCCAAGCAGCTGCCGGCTTCAC 1006

QY 948 CACCTTCACCATCGCCGACAGATCACCTCTCAAGGCTGCCTGCCTGGACAT 1001

Db 1007 CACCTTCACCATCGCCGACAGATCACCTCTCAAGGCTGCCTGCCTGGACAT 1060

RESULT 15

AK098172

LOCUS

DEFINITION Homo sapiens CDNA FLJ40853 f1s, clone TRACH2015486, highly similar
to RETINOIC ACID RECEPTOR ALPHA.

ACCESSION AK098172

VERSION AK098172.1

KEYWORDS GI:21758129

SOURCE Homo sapiens trachea CDNA to mRNA, clone_lib:TRACH2

clone:TRACH2015486.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Suzuki, O., Sasaki, N., Aotsuka, S., Shoji, T., Ichihara, T.,

Shiohata, N., Matsumoto, K., Hirano, M., Sano, S., Nomura, R.,

Yoshikawa, Y., Matsumura, Y., Moriya, S., Chiba, E., Momiya, H.,

Onogawa, S., Kaeriyama, S., Satoh, N., Matsunawa, H., Takahashi, E.,

Kataoka, R., Kuga, N., Kuroda, A., Satoh, I., Kamata, K., Takami, S.,

Terashima, Y., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T.,

Sato, H., Ota, Y., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,

Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,

Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,

Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,

Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,

Nagahari, K., Masuno, Y., Nagai, K. and Isogai, T.

NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 2008)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; CDNA full insert sequencing:

Research Association for Biotechnology (RAB); CDNA library

construction: Helix Research Institute (HRI) (supported by Japan

Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

HRI, and Biotechnology Center, National Institute of Technology and

Evaluation; clone selection for full insert sequencing: HRI and

RAB; annotation: HRI and RAB.
Location/Qualifiers
1..2008
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TRACH2015486"
/tissue_type="trachea"
/clone_lib="TRACH2"
/note="cloning vector: pME18SFL3"
BASE COUNT : 405 a 669 c 576 g 358 t
ORIGIN

Query Match 17.4%; Score 174; DB 9; Length 2008;
Best Local Similarity 100.0%; Pred. No. 2.4e-27;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 828 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTTGACCTCTGGGACAAAGTTTCAGTGA 887
|||||
Db 680 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTTGACCTCTGGGACAAAGTTTCAGTGA 739
|||||

Oy 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCCGCCAAGCAGCTGCCCGGCTTCAC 947
|||||
Db 740 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCCGCCAAGCAGCTGCCCGGCTTCAC 799
|||||

Oy 948 CACCCTCACCATCGCCGACCATCACCCTCCTCAAGGCTGCCTGGACAT 1001
|||||
Db 800 CACCCTCACCATCGCCGACCATCACCCTCCTCAAGGCTGCCTGGACAT 853
|||||

Search completed: March 30, 2003, 13:00:12
Job time : 2850.53 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 07:57:38 ; Search time 196.299 Seconds
(without alignments)
11483.757 Million cell updates/sec

Title: US-09-691-220-3_COPY_10000_11000

Perfect score: 1001

Sequence: 1 cccagggagactgcagctgg.....aaggctgcctgcctggacat 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1001	100.0	20512	24	Genomic DNA encodi
2	174	17.4	1481	17	RAR-alpha RNA sequ
3	174	17.4	1481	17	RAR-alpha DNA sequ
4	174	17.4	1692	23	DNA encoding novel
5	174	17.4	1944	23	DNA encoding novel
6	174	17.4	1944	23	DNA encoding novel
7	174	17.4	2073	23	DNA encoding novel
8	174	17.4	2086	24	CDNA encoding huma
9	174	17.4	2457	23	DNA encoding novel

10	174	17.4	2658	16	AAQ81477	
11	174	17.4	2907	24	ABK84517	
12	174	17.4	2907	24	ABL65214	
13	174	17.4	2928	13	AAQ29338	
14	174	17.4	2928	20	AAV64991	
15	174	17.4	2940	10	AA90124	
16	174	17.4	2940	16	AAO81476	
17	174	17.4	3036	13	AAQ29334	
18	174	17.4	3036	19	AAV20474	
19	174	17.4	3036	20	AAV64990	
20	174	17.4	3052	23	AA83047	
21	174	17.4	3511	17	AAT33246	
22	174	17.4	3511	17	AAT33259	
23	174	17.4	3511	21	AAA38656	
24	156.6	15.6	2930	23	AA83048	
c	25	156	301	22	AA79992	
26	140.8	14.1	456	23	AA83044	
27	117.6	11.7	1375	24	ABK37492	
28	117.6	11.7	1399	24	ABK37493	
29	117.6	11.7	8887	24	ABK37494	
30	117.6	11.7	9151	24	ABK37495	
31	112.8	11.3	1351	16	AAQ79935	
32	112.8	11.3	1912	12	AAO10388	
33	112.8	11.3	1913	16	AAT05202	
34	112.8	11.3	2521	13	AAQ20542	
35	112.8	11.3	2740	13	AAQ20541	
36	112.8	11.3	2989	10	AA900093	
37	112.8	11.3	2992	16	AAQ79934	
38	112.8	11.3	2992	17	AAT06491	
39	111.2	11.1	2992	15	AAQ65572	
40	109.2	10.9	1576	12	AAQ10226	
41	109.2	10.9	1903	12	AAQ10389	
42	109.2	10.9	1903	16	AAT05196	
43	107.2	10.7	1416	23	AA86580	
c	44	102.4	10.2	482	23	AA81232
45	102.4	10.2	5838	22	AAH98429	

ALIGNMENTS

RESULT 1
AAL38339
ID AAL38339 standard; DNA; 20512 BP.
XX AC AAL38339;
XX 15-AUG-2002 (first entry)
XX Genomic DNA encoding the human nuclear hormone receptor protein.
XX Human nuclear hormone receptor; therapeutic agent; gene therapy;
KW Immune response; chromosome 17; single nucleotide polymorphism;
KW gene; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
FH CDS 1684..15208
FT /*tag= a
FT /*product= "Human nuclear hormone receptor protein"
FT exon 1684..1846
FT /*tag= b
FT /*number= 1
FT intron 1847..6808
FT /*tag= c
FT exon 6809..6957
FT /*tag= d
FT intron 6958..8277
FT /*tag= e
FT /*number= 2

```
FT exon      8278..8419
FT /*tag= f
FT /number= 3
FT intron    8420..10406
FT /*tag= g
FT exon      10407..10567
FT /*tag= h
FT intron    10568..10827
FT /*tag= i
FT exon      10828..11004
FT /*tag= j
FT intron    11005..12798
FT /*tag= k
FT exon      12799..13003
FT /*tag= l
FT intron    13004..13759
FT /*tag= m
FT exon      13760..13918
FT /*tag= n
FT intron    13919..14505
FT /*tag= o
FT exon      14506..14658
FT /*tag= p
FT intron    14659..15143
FT /*tag= q
FT exon      15144..15208
FT /*tag= r
FT allele    replace(4084,C)
FT allele    replace(6482,A)
FT allele    replace(8066,G)
FT allele    replace(8699,C)
FT allele    replace(12897,T)
FT allele    replace(14442,C)
FT /*tag= x
XX WO200231146-A2.
PN 18-APR-2002.
XX
XX 05-OCT-2001; 2001WO-US31095.
XX
XX 11-OCT-2000; 2000US-239117P.
XX 19-OCT-2000; 2000US-0691220.
XX
XX (PEKE ) PE CORP NY.
XX
XX Wei M, Ye J, Yan C, Ketchum KA, Di Francesco V, Beasley EM;
XX WPI: 2002-426282/45.
XX P-PSDB; AAO21489.
XX
XX New human nuclear hormone receptor proteins and nucleic acids, useful
XX as models or targets for developing human therapeutic targets, and in
XX identifying therapeutic proteins and modulators of nuclear hormone
XX receptor expression
```

PS Claim 1; Fig 3; 73pp; English.

XX The invention relates to an isolated peptide of a novel human nuclear hormone receptor with a fully defined sequence of 457 amino acids given in the specification. The novel human nuclear hormone receptor peptides and nucleic acids encoding them can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate nuclear hormone receptor activity in cells and tissues that express the nuclear hormone receptor. The nucleic acids may be used as a query sequence to perform searches against sequence databases to identify family members or related sequences, as probes or primers, to construct recombinant vectors, to identify compounds that modulate nuclear hormone receptor nucleic acid expression, in gene therapy, and as antisense constructs to control nuclear hormone receptor gene expression in cells, tissues or organisms. The polypeptides can be used to raise antibodies or to elicit an immune response, as a reagent in assays designed to determine protein levels in biological fluids, as markers for tissues in which a corresponding protein is expressed, to identify a binding partner/ligand to develop a system for the identification of inhibitors of the binding reaction, in drug screening assays, and to identify compounds that modulate protein activity. This polynucleotide sequence represents the genomic DNA encoding the human nuclear hormone receptor protein of the invention.

XX SQ Sequence 20512 BP; 3591 A; 5987 C; 6281 G; 4541 T; 112 other;

Query Match 100.0%; Score 1001; DB 24; Length 20512;
Best Local Similarity 100.0%; Pred. No. 5.3e-223;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGGGAGACTGCAGCTGGGAGGGCTGGGTGAGTGGAGCGGGAGAGACTTCCTGG 60
Db 10000 CCCAGGGAGACTGCAGCTGGGAGGGCTGGGTGAGTGGAGCGGGAGAGACTTCCTGG 10059
QY 61 GGAAGAGGAGGAGCAGACCTAGGAGGSCACCGCTGAGTGGAGTGGAGTGGAGTGA 120
Db 10060 GGAAGAGGAGGAGCAGACCTAGGAGGSCACCGCTGAGTGGAGTGGAGTGA 10119
QY 121 CGGTGGGGATAGCATCGGCTGGCTATGGGGTGGGTGGGGGTGGTGTGAGGGGCCAC 180
Db 10120 CGGTGGGGATAGCATCGGCTGGCTATGGGGTGGGTGGGGGTGGTGTGAGGGGCCAC 10179
QY 181 AGCTGTGCTCATGGGCTTCCTGGGCGACAACCTTGATGTGGGTGGGTGGGCATGGAGG 240
Db 10180 AGCTGTGCTCATGGGCTTCCTGGGCGACAACCTTGATGTGGGTGGGTGGGCATGGAGG 10239
QY 241 GCTGGAGTGGCTGGCAATGCCTTGCTGCCGCTGAACGCGTCTGTGTGCGGTGCTTAC 300
Db 10240 GCTGGAGTGGCTGGCAATGCCTTGCTGCCGCTGAACGCGTCTGTGTGCGGTGCTTAC 10299
QY 301 AAGCCTGGGTGACCTCTCAGCAGCTGGCAGCTCTCTGTCTAGGCTGGGGGTGGAGGAGC 360
Db 10300 AAGCCTGGGTGACCTCTCAGCAGCTGGCAGCTCTCTGTCTAGGCTGGGGGTGGAGGAGC 10359
QY 361 CCTGAGCAGCTGCAGCTGCCCTCTTAACCCCTCTGCCCTCCACAGCTGTGAGAAACA 420
Db 10360 CCTGAGCAGCTGCAGCTGCCCTCTTAACCCCTCTGCCCTCCACAGCTGTGAGAAACA 10419
QY 421 CCGAAACAAGAAGAAGAGGAGTGGCCCAAGCCGAGTGCCTCTGAGAGCTACACGCTGAC 480
Db 10420 CCGAAACAAGAAGAAGAGGAGTGGCCCAAGCCGAGTGCCTCTGAGAGCTACACGCTGAC 10479
QY 481 GCCGGAGTGGGGAGCTCATTTGAGAAGGTGCGAAAGCGCACCCAGGAAACCTTCCTGTC 540
Db 10480 GCCGGAGTGGGGAGCTCATTTGAGAAGGTGCGCAAAAGCGCACCCAGGAAACCTTCCTGTC 10539
QY 541 CCTCTGCCAGCTGGGCAAAATACACTACGCTATGCTTTCCCGGCTGCAAGGTGGGAT 600
Db 10540 CCTCTGCCAGCTGGGCAAAATACACTACGCTATGCTTTCCCGGCTGCAAGGTGGGAT 10599
QY 601 TTCCCGAGGSCACAGGCGCCAGGATGGCCCTCTCAGGACACCCCTCTCTTGTGCGCAGCA 660
|||||

Db	10600	TTGCCAGGGCCACAGAGCCAGAGTGGGCCCTCTCAGGCACCCCTTCTTGTCACAGGCA	10555
Qy	661	AGATCTCTGCGTCTCTTCCCTTTCCCTCTCTTTCTGCTGCTGCTCTTCCCAAG	720
Db	10660	AGATCTCTGCGTCTCTTCCCTTCCCTCTCTTCTCCCTCTCTGCTGCTTCCCAAG	10719
Qy	721	AGCTCCCAAGAGTGAAGGCTGGGTAGAGGCGAGCCCTGTGGGGCTGGAGCCAGGCTGA	780
Db	10720	AGCTCCCAAGAGTGAAGGCTGGGTAGAGGCGAGCCCTGTGGGGCTGGAGCCAGGCTGA	10779
Qy	781	GAAGGGTGCCATGGAGAGAAAGCCCTCACTCTCCCTCCCCAGACAACAGCTCA	840
Db	10780	GAAGGGTGCCATGGAGAGAAAGCCCTCACTCTCCCTCCCCAGACAACAGCTCA	10839
Qy	841	GAACAACGTGCTCTCTGGACATTGACCTCTGGGACAAAGTTTCACTGAACCTCCACCAAG	900
Db	10840	GAACAACGTGCTCTCTGGACATTGACCTCTGGGACAAAGTTTCACTGAACCTCCACCAAG	10899
Qy	901	TGCATCATTAAGACTGTGGAGTTTCCGCAAGCAGCTGCCCGCTTCAACACCTCACCATC	960
Db	10900	TGCATCATTAAGACTGTGGAGTTTCCGCAAGCAGCTGCCCGCTTCAACACCTCACCATC	10959
Qy	961	GCCACACAGATCACCTCCTCAAGGCTGCCTGGACAT	1001
Db	10960	GCCACACAGATCACCTCCTCAAGGCTGCCTGGACAT	11000
RESULT 2			
AAT33247			
ID	AAT33247 standard; RNA; 1481 BP.		
XX			
AC	AAT33247;		
XX			
DT	09-OCT-1996 (first entry)		
XX			
XX	RAR-alpha RNA sequence.		
DE			
KW	Hammerhead ribozyme; acute promyelocytic leukaemia; APL;		
KW	lymphoma; therapy; PML-RAR-alpha; retinoic acid receptor;		
KW	external guide sequence; EGS; antisense; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	W09618733-A2.		
XX			
PD	20-JUN-1996.		
XX			
PF	14-DEC-1995; 95WO-US16451.		
XX			
PR	14-DEC-1994; 94US-0354956.		
XX			
PA	(INNO-) INNOVIR LAB INC.		
XX			
PI	George ST, Goldberg AR, Pace U;		
XX			
DR	WPI; 1996-300650/30.		
XX			
PT	RNA construct(s) including ribozyme(s) and antisense		
PT	oligo:nucleotide(s) - for the inactivation of RNA associated with,		
PT	e.g. promyelocytic leukaemia or follicular lymphoma		
XX			
Example 1; Page 55-56; 81pp; English.			
XX			
CC	Retinoic acid receptor (RAR-alpha) RNA (AAT33247) is a substrate for		
CC	ribozymes; external guide sequences and antisense oligonucleotides		
CC	(see also AAT33245 and AAT33248-58) used for the treatment of acute		
CC	promyelocytic leukaemia (APL). APL is characterised by a balanced,		
CC	reciprocal translocation between the long arms of chromosomes 15		
CC	and 17, resulting in a fusion of the retinoic acid receptor gene		
CC	(see also AAT33260) and a gene for a putative transcription factor,		
CC	(PML. The ribozymes etc. specifically cleave the PML-RAR-alpha fusion		
CC	mRNA but not wild-type RAR-alpha mRNA.		

SQ Sequence 1481 BP; 308 A; 488 C; 498 G; 260 U; 0 other;

Query Match . 17.4%; Score 174; DB 17; Length 1481;
Best Local Similarity 79.9%; Pred. No. 6.1e-31;
Matches 139; Conservative 35; Mismatches 0; Indels 0; Gaps 0

Qy 828 GAACAACAGCTCAGAACAACGTTGTCTCTTGACATTGACCTTCTGGGCAAAAGTTTCAGTGA 887
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 597 GAACAACAGCUCAGAACACGUGUCUCUCUGGACAUAGACCUCUGGGACAGUUCAGUGA 656
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 888 ACTCTCCACCACAAAGTGCATCATTAAGACTGTGTGGAGTTCCGCCAAGCAGCTGCCCGGTTTCAC 947
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 657 ACUCUCCACCAAGUGCAUCAUAUAGACUGUGGAGUUCCCAAGCAGCUGCCCCGGCUDCAC 716
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 948 CACCTCACCATCGCCGACCAAGACACCTCTCTCAAGCTGCCTCGCTGCCTGGACAT 1001
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 717 CACCCUACCAUCGCGGACCAAGACCCUCCUCAAAGGUGCCUGCCUGGACAU 770

RESULT 3
AAT33260
ID AAT33260 standard; cDNA; 1481 BP.
XX AAT33260;
XX AC
XX XX
DT 09-OCT-1996 (first entry)
XX RAR-alpha DNA sequence.
XX
KW Hammerhead ribozyme; acute promyelocytic leukaemia; APL;
KW lymphoma; therapy; PML-RAR-alpha; retinoic acid receptor;
KW external guide sequence; EGS; antisense; ss.
XX
OS Homo sapiens.
XX
PN W09618733-A2.
XX
PD 20-JUN-1996.
XX
XX 14-DEC-1995; 95WO-US16451.
PF
XX
PR 14-DEC-1994; 94US-0354956.
XX
PA (INNO-) INNOVIR LAB INC.
XX
PI George ST, Goldberg AR, Pace U;
XX
DR WPI; 1996-300650/30..
XX
XX RNA construct(s) including ribozyme(s) and antisense
PT oligo:nucleotide(s) - for the inactivation of RNA associated with,
PT e.g. promyelocytic leukaemia or follicular lymphoma
XX
PS Example 1; Page 52-53; 81pp; English.
XX
XX Acute promyelocytic leukaemia (APL) is associated with a
CC translocation between the long arms of chromosomes 15 and 17,
CC resulting in the fusion between the retinoic acid receptor gene
CC (RAR-alpha, AAT33260) and PML, a putative transcription factor gene.
CC The fusion product RNA (see also AAT33246) junction region (AAT33251)
CC is the target for ribozymes (AAT33245, AAT33248, AAT33250 and AAT33252),
CC external guide sequences (AAT33253-54) and antisense constructs
CC (AAT33255-57) that specifically cleave the PML-RAR-alpha fusion
XX mRNA but not wild-type RAR-alpha mRNA (see also AAT33247).
XX
SQ Sequence 1481 BP; 309 A; 487 C; 425 G; 257 T; 3 U; 0 other;

Query Match 17.4%; Score 174; DB 17; Length 1481;
Best Local Similarity 100.0%; Pred. No. 6.1e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 828 GACAACAGCTCAGAACACGTTGTCTCTTGACATTGACCTCTGGGCAAAAGTTTCAGTGA 887
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 597 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 656
QY 888 ACTCTCACCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGACAGCTGCCCGGCTTCAC 947
Db 657 ACTCTCACCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGACAGCTGCCCGGCTTCAC 716
QY 948 CACCTCACCATCCCGCAGCAGATCACCTCTCAAGCTGCCCTGGGACAT 1001
Db 717 CACCTCACCATCCCGCAGCAGATCACCTCTCAAGCTGCCCTGGGACAT 770

RESULT 4
AAS83808
ID AAS83808 standard; cdna; 1692 BP.
XX AAS83808;
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #19612.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.

XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG19621.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 1; SEQ ID No 19612; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1692 BP; 414 A; 470 C; 492 G; 316 T; 0 other;

Query Match 17.4%; Score 174; DB 23; Length 1692;
Best Local Similarity 100.0%; Pred. No. 6.3e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887
Db 933 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 992
QY 888 ACTCTCACCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGACAGCTGCCCGGCTTCAC 947
Db 993 ACTCTCACCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGACAGCTGCCCGGCTTCAC 1052
QY 948 CACCTCACCATCCCGCAGCAGATCACCTCTCAAGCTGCCCTGGGACAT 1001
Db 1053 CACCTCACCATCCCGCAGCAGATCACCTCTCAAGCTGCCCTGGGACAT 1106

RESULT 5
AAS83045
ID AAS83045 standard; cdna; 1944 BP.

XX AAS83045;
XX
DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #18849.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.

XX WO200175067-A2.
PN 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.

DR P-PSDB; ABG18858.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 1; SEQ ID No 18849; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 1944 BP; 427 A; 598 C; 557 G; 362 T; 0 other;
Query Match 17.4%; Score 174; DB 23; Length 1944;
Best Local Similarity 100.0%; Pred. No. 6.5e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGTCTCTGGACATTCACCTCGGACAAAGTTTCAGTGA 887
DB 887 GAACAACAGCTCAGAACACGTGTCTCTGGACATTCACCTCGGACAAAGTTTCAGTGA 946
QY 888 ACTCTCCACCAAGTCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 947
DB 947 ACTCTCCACCAAGTCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 1006
QY 948 CACCTTCACCATCGCGACAGATCACCTCTCAAGGCTGCCTGGACAT 1001
DB 1007 CACCTTCACCATCGCGACAGATCACCTCTCAAGGCTGCCTGGACAT 1060

RESULT 6
AAS85136
ID AAS85136 standard; cDNA; 1944 BP.
XX AC AAS85136;
XX AC AAS85136;
DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #20940.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG20949.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX Claim 1; SEQ ID NO 20940; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 1944 BP; 427 A; 598 C; 557 G; 362 T; 0 other;
Query Match 17.4%; Score 174; DB 23; Length 1944;
Best Local Similarity 100.0%; Pred. No. 6.5e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGTCTCTGGACATTCACCTCTGGACAAAGTTTCAGTGA 887
DB 887 GAACAACAGCTCAGAACACGTGTCTCTGGACATTCACCTCTGGACAAAGTTTCAGTGA 946
QY 888 ACTCTCCACCAAGTCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 947
DB 947 ACTCTCCACCAAGTCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 1006
QY 948 CACCTTCACCATCGCGACAGATCACCTCTCAAGGCTGCCTGGACAT 1001
DB 1007 CACCTTCACCATCGCGACAGATCACCTCTCAAGGCTGCCTGGACAT 1060

RESULT 7
AAS85135
ID AAS85135 standard; cDNA; 2073 BP.
XX AC AAS85135;
XX AC AAS85135;
DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #20939.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG20948.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX Claim 1; SEQ ID NO 20939; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS94197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 2073 BP; 465 A; 612 C; 605 G; 391 T; 0 other;

Query Match 17.4%; Score 174; DB 23; Length 2073;
Best Local Similarity 100.0%; Pred. No. 6.6e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAGTGA 887

Db 1016 GAACAACAGCTCAGAACACGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAGTGA 1075

QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 947

Db 1076 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 1135

QY 948 CACCTCACCATCGCCGACAGATCACCCTCTCAAGCTGCCCTGGGACAT 1001

Db 1136 CACCTCACCATCGCCGACAGATCACCCTCTCAAGCTGCCCTGGGACAT 1189

RESULT 8

AAL38338

ID AAL38338 standard; cDNA; 2086 BP.

XX AAL38338;

XX 15-AUG-2002 (first entry)

DE cDNA encoding human nuclear hormone receptor protein.

XX Human nuclear hormone receptor; therapeutic agent; gene therapy;

XX Immune response; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 127..1500

FT FT /*tag= a

FT FT /product= "Human nuclear hormone receptor protein"

XX WO200231146-A2.

XX 18-APR-2002.

XX 05-OCT-2001; 2001WO-US31095.

XX 11-OCT-2000; 2000US-239117P.

PR 19-OCT-2000; 2000US-0691220.

XX (PEKE) PE CORP NY.

XX Wei M, Ye J, Yan C, Ketchum KA, Di Francesco V, Beasley EM;

XX WPI; 2002-426282/45.

DR P-PSDB; AAO21489.

XX New human nuclear hormone receptor proteins and nucleic acids, useful

XX as models or targets for developing human therapeutic targets, and in

PT identifying therapeutic proteins and modulators of nuclear hormone

XX receptor expression

PS claim 1; Fig 1; 73pp; English.

XX The invention relates to an isolated peptide of a novel human nuclear
CC hormone receptor with a fully defined sequence of 457 amino acids given
CC in the specification. The novel human nuclear hormone receptor peptides
CC and nucleic acids encoding them can be used as models for the development
CC of human therapeutic targets, aid in the identification of therapeutic
CC proteins, and serve as targets for the development of human therapeutic
CC agents that modulate nuclear hormone receptor activity in cells and
CC tissues that express the nuclear hormone receptor. The nucleic acids may
CC be used as a query sequence to perform searches against sequence
CC databases to identify family members or related sequences, as probes or
CC primers, to construct recombinant vectors, to identify compounds that
CC modulate nuclear hormone receptor nucleic acid expression, in gene
CC therapy, and as antisense constructs to control nuclear hormone receptor
CC gene expression in cells, tissues or organisms. The polypeptides can be
CC used to raise antibodies or to elicit an immune response, as a reagent in
CC assays designed to determine protein levels in biological fluids, as
CC markers for tissues in which a corresponding protein is expressed, to
CC identify a binding partner/ligand to develop a system for the
CC identification of inhibitors of the binding reaction, in drug screening
CC assays, and to identify compounds that modulate protein activity. This
CC polynucleotide sequence represents the cDNA encoding the human nuclear
CC hormone receptor protein of the invention.

XX SQ Sequence 2086 BP; 437 A; 685 C; 583 G; 381 T; 0 other;

Query Match 17.4%; Score 174; DB 24; Length 2086;

Best Local Similarity 100.0%; Pred. No. 6.6e-31;

Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAGTGA 887

Db 741 GAACAACAGCTCAGAACACGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAGTGA 800

QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 947

Db 801 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 860

QY 948 CACCTCACCATCGCCGACAGATCACCCTCTCAAGCTGCCCTGGGACAT 1001

Db 861 CACCTCACCATCGCCGACAGATCACCCTCTCAAGCTGCCCTGGGACAT 914

RESULT 9

AAS83049

ID AAS83049 standard; cDNA; 2457 BP.

XX AAS83049;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #18853.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
DR P-PSDB; ABCI8862.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1: SEQ ID No 18853; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2457 BP; 577 A; 687 C; 700 G; 493 T; 0 other;
Query Match 17.4%; Score 174; DB 23; Length 2457;
Best Local Similarity 100.0%; Pred. No. 6.8e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 887
DB 1400 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 1459
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 947
DB 1460 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 1519
QY 948 CACCCTCACCATCGCGGACCAAGATCACCTCTCTCAAGGCTGCCTGGACAT 1001
DB 1520 CACCCTCACCATCGCGGACCAAGATCACCTCTCTCAAGGCTGCCTGGACAT 1573
RESULT 10
AAQ81477
ID AAQ81477 standard; cDNA to mRNA; 2658 BP.
XX
AC AAQ81477;
XX
DT 01-SEP-1995 (first entry)
XX
DE RAR-alpha-403 dominant negative.
XX
KW RAR-alpha; retinoic acid receptor alpha; hematopoietic; stem cell;
KW differentiation; dominant negative; retrovirus; vector; neutrophil;
KW monocyte; mast cell; basophil; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 103..1312
FT /*tag= a
FT /note= "deletion of bases 1311-1596 of RAR-alpha

CDNA (AAQ81476) and insertion of stop codon
TAG at deletion site (1312-1314)."

WO9504143-A.
XX
XX 09-FEB-1995.
XX
XX 28-JUL-1994; 94WO-US08450.
XX
XX 28-JUL-1993; 93US-0099242.
XX
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
XX Collins SJ, Tsai S;
XX
XX WPI: 1995-082227/11.
XX P-PSDB; AAR68024.
XX
XX New hematopoietic stem cell lines with specific differentiation
XX properties - made by transfecting stem cells with nucleic acid
XX encoding dominant negative suppressor of the retinoic acid
XX receptor alpha, useful e.g. for hematopoietic reconstitution
XX
XX Disclosure; Page 58-59; 100pp; English.
XX
XX An RAR-alpha cDNA (given in AAQ81477) contained a truncation of
XX sequences coding for the C-terminal 59 amino acids and part the
XX 3' UTR of the wild-type human sequence (AAQ81476). This truncated
XX cDNA, designated RAR-alpha-403, encoded a protein (AAR68024)
XX containing the N-terminus, DNA-binding domain and part of the
XX hormone-binding domain of RAR-alpha (AAR68023). RAR-alpha-403 has
XX negative suppressor activity. Introduction into multipotent IL-3
XX dependent FDCP mix A4 cells (murine hematopoietic) caused a
XX switch from neutrophil/monocyte differentiation to basophil/mast
XX cells (30-50%), with only 2% neutrophils.
XX
SQ Sequence 2658 BP; 550 A; 889 C; 711 G; 508 T; 0 other;
Query Match 17.4%; Score 174; DB 16; Length 2658;
Best Local Similarity 100.0%; Pred. No. 7e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 887
DB 732 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 791
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 947
DB 792 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 851
QY 948 CACCCTCACCATCGCGGACCAAGATCACCTCTCTCAAGGCTGCCTGGACAT 1001
DB 852 CACCCTCACCATCGCGGACCAAGATCACCTCTCTCAAGGCTGCCTGGACAT 905
RESULT 11
ABK84517
ID ABK84517 standard; cDNA; 2907 BP.
XX
AC ABK84517;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #1088.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.
XX WO200228999-A2.
XX 11-APR-2002.
XX 03-OCT-2001; 2001WO-US30821.
XX 03-OCT-2000; 2000US-237189P.
XX (GENE-) GENE LOGIC INC.
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX Detecting granulocyte activation by detecting differential expression
XX of genes associated with granulocyte activation, which serves as
XX diagnostic markers that is useful for monitoring disease states and
XX drug toxicity -
XX Claim 1; SEQ ID No 1088; 114pp; English.
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing
XX the expression level to an expression level in an unactivated
XX GC, where differential expression of Gs is indicative of GCA.
XX Also included are modulating (M2) GCA by contacting GC with an agent
XX that alters the expression of at least one gene in Gs; (2) screening (M3)
XX for an agent capable of modulating GCA or an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease using the
XX gene expression profile; (3) detecting (M4) an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease, by detecting the
XX level of expression in a sample of the tissue of gene(s) from Gs, where
XX the level of expression of the gene is indicative of inflammation;
XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,
XX an allergic response in a subject, exposure of a subject to a pathogen
XX or sterile inflammatory disease, by contacting a tissue having
XX inflammation with an agent that modulates the expression of gene(s)
XX from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
XX modulating GCA; M3 is useful for screening an agent capable of modulating
XX GCA preferably in an inflammation in a tissue; M4 is useful for
XX detecting an inflammation (especially chronic) in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
XX glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
XX reperfusion injury, AIDS, adult respiratory distress syndrome,
XX inflammatory bowel disease, Crohn's disease, ulcerative colitis,
XX periodontal disease; also bacterial infection, viral infection,
XX parasitic infection, protozoal infection, fungal infection and M5 is
XX useful for treating one of the above conditions. The present
XX sequence represents a gene differentially expressed in granulocytes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 2907 BP; 561 A; 1011 C; 789 G; 546 T; 0 other;
Query Match 17.4%; Score 174; DB 24; Length 2907;
Best Local Similarity 100.0%; Pred. No. 7.1e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGACAAGCTGTCTCTGGACATTCACCTCTGGCAGACAGTTCACTGA 887
DB 732 GAACAACAGCTCAGACAAGCTGTCTCTGGACATTCACCTCTGGCAGACAGTTCACTGA 791
QY 888 ACTCTCCACCAAGTGCATCATTAAGAGCTGTGGAGTTCCCAAGCAGCTGCCCGCTTCAC 947
|||||

Db 792 ACTCTCCACCAAGTGCATCATTAAGAGCTGTGGAGTTCCCAAGCAGCTGCCCGCTTCAC 851
QY 948 CACCTCTACCATCGCCGACAGATCACCTCTCTAAGGCTGCCTCGCTGGACAT 1001
|||||
Db 852 CACCTCTACCATCGCCGACAGATCACCTCTCTAAGGCTGCCTCGCTGGACAT 905
RESULT 12
ABL65214
ID ABL65214 standard; DNA; 2907 BP.
XX AC ABL65214;
XX
XX 15-MAY-2002 (first entry)
XX Lung cancer related gene sequence SEQ ID NO:3551.
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytosstatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
XX gene; ds.
XX Homo sapiens.
XX OS
XX WO200194629-A2.
XX PN
XX 13-DEC-2001.
XX PD
XX 30-MAY-2001; 2001WO-US10838.
XX 05-JUN-2000; 2000US-209473P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-233133P.
XX 20-SEP-2000; 2000US-233617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 22-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234509P.
XX 22-SEP-2000; 2000US-234567P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-234924P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 25-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 26-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
XX 27-SEP-2000; 2000US-235840P.
XX 27-SEP-2000; 2000US-235863P.
XX 28-SEP-2000; 2000US-236028P.
XX 28-SEP-2000; 2000US-236032P.
XX 28-SEP-2000; 2000US-236033P.
XX 28-SEP-2000; 2000US-236034P.
XX 28-SEP-2000; 2000US-236109P.
XX 28-SEP-2000; 2000US-236111P.
XX 29-SEP-2000; 2000US-236842P.
XX 29-SEP-2000; 2000US-236891P.
XX 02-OCT-2000; 2000US-237172P.
XX 02-OCT-2000; 2000US-237173P.
XX 02-OCT-2000; 2000US-237278P.
XX 02-OCT-2000; 2000US-237294P.
XX 02-OCT-2000; 2000US-237295P.
XX 02-OCT-2000; 2000US-237316P.
XX 03-OCT-2000; 2000US-237425P.
XX 03-OCT-2000; 2000US-237598P.
XX 03-OCT-2000; 2000US-237604P.
XX 03-OCT-2000; 2000US-237608P.
XX 03-OCT-2000; 2000US-237608P.
XX 01-NOV-2000; 2000US-244867P.
XX 01-NOV-2000; 2000US-245084P.
XX

PA (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI: 2002-188264/24.
DR
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1: SEQ ID 3551; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 2907 BP; 561 A; 1011 C; 789 G; 546 T; 0 other;

Query Match 17.4%; Score 174; DB 24; Length 2907;
Best Local Similarity 100.0%; Pred. No. 7.1e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTCAAGTGA 887
DB 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTCAAGTGA 791

OY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGCTTCAC 947
DB 792 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGCTTCAC 851

OY 948 CACCCTCACCATTCCGCGACCATCACCCTCTCTCAAGGCTGCTGCTGGACAT 1001
DB 852 CACCCTCACCATTCCGCGACCATCACCCTCTCTCAAGGCTGCTGCTGGACAT 905

RESULT 13
AAQ29338
ID AAQ29338 standard; cDNA; 2928 BP.
XX
AC AAQ29338;
XX
DT 09-MAR-1993 (first entry)
XX
DE RAR-alpha gene.
XX
KW Retinolic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;
KW APL; translocation; chromosome 17; chromosome 15; PCR; primer;
KW [t(15;17)(q21;q11-22)]; breakpoint; polymerase chain reaction; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 103..1488
FT /*tag= a
XX
PN W09216660-A.
XX
PD 01-OCT-1992.

XX 23-MAR-1992; 92WO-US02320.
XX
XX 22-MAR-1991; 91US-0673838.
PR 22-MAR-1991; 91US-0675084.
XX
XX (SLOK) SLOAN KETTERING INST CANCER.
XX
XX Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;
PI Warrell RP;
XX
XX WPI: 1992-349240/42.
DR P-PSDB; AAR27534.
XX
XX Marker for acute promyelocytic leukaemia and other neoplasias -
PT comprising nucleic acid and encoded abnormal retinoic acid
PT receptor-alpha receptor
XX
XX Disclosure; Page 43-46; 84pp; English.
XX
CC The sequence given represents the nucleic acid sequence of the
CC retinoic acid receptor (RAR)-alpha gene. This gene is disrupted in
CC a translocation of a portion of the long arm of chromosome 17 onto
CC the long arm of chromosome 15 [t(15;17)(q21;q11-22)]. This causes a
CC fusion between RAR-alpha and myl which is characteristic of acute
CC promyelocytic leukemia (APL). The breakpoint region has been cloned
CC and it has been shown that DNA rearrangements are clustered in the
CC region of the first intron of RAR-alpha. This sequence was isolated
CC by polymerase chain reaction (PCR). The primers used for amplification
CC of this sequence can also be used to amplify the translocated region.
XX
XX Sequence 2928 BP; 582 A; 1012 C; 789 G; 545 T; 0 other;

Query Match 17.4%; Score 174; DB 13; Length 2928;
Best Local Similarity 100.0%; Pred. No. 7.1e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTCAAGTGA 887
DB 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTCAAGTGA 791

OY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGCTTCAC 947
DB 792 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGCTTCAC 851

OY 948 CACCCTCACCATTCCGCGACCATCACCCTCTCTCAAGGCTGCTGCTGGACAT 1001
DB 852 CACCCTCACCATTCCGCGACCATCACCCTCTCTCAAGGCTGCTGCTGGACAT 905

RESULT 14
AAV64991
ID AAV64991 standard; cDNA; 2928 BP.
XX
AC AAV64991;
XX
DT 05-FEB-1999 (first entry)
XX
DE Human RAR-alpha cDNA.
XX
KW Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;
KW acute promyelocytic leukaemia; APL; t(15;17); translocation;
KW treatment; all-trans retinoic acid; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 103..1491
FT /*tag= a
FT /product= "RAR-alpha"
FT /note= "retinoic acid receptor-alpha"
XX
PN US5843642-A.

```
XX 01-DEC-1998.
XX PD
XX PF
XX 21-JUL-1993; 93US-0095728.
XX 21-JUL-1993; 93US-0095728.
PR 22-MAR-1991; 91US-0673838.
PR 22-MAR-1991; 91US-0675084.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX PI
XX Dmitrovsky E, Frankel S, Miller WH, Warrell RP;
XX WPI: 1999-044563/04.
XX DR P-PSDB; AAW81964.
XX
XX Diagnosis of acute promyelocytic leukaemia - by detecting nucleic
PT acid encoding abnormal retinoic acid receptor-alpha
XX PS
XX Disclosure; Column 29-34; 38pp; English.
XX
XX This sequence encodes the human retinoic acid receptor alpha, RAR-alpha
CC which is used in a method for identifying a subject with acute
CC promyelocytic leukaemia (APL) resulting from a t(15;17) translocation who
CC will respond to treatment with all-trans retinoic acid. The protein can
CC also be used to identify a subject with indications of APL who will not
CC respond to treatment with all-trans retinoic acid.
XX
XX Sequence 2928 BP; 582 A; 1011 C; 789 G; 546 T; 0 other;
SQ
Query Match 17.4%; Score 174; DB 20; Length 2928;
Best Local Similarity 100.0%; Pred. No. 7.1e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGACAACGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTCACTGA 887
DB 732 GAACAACAGCTCAGACAACGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTCACTGA 791
QY 888 ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGCTTCAC 947
DB 792 ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGCTTCAC 851
QY 948 CACCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGGGACAT 1001
DB 852 CACCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGGGACAT 905
```

```
RESULT 15
AAN90124
ID AAN90124 standard; DNA; 2940 BP.
XX AC
XX AAN90124;
XX
DT 01-NOV-1989 (first entry)
XX
DE DNA of clone phrAR1.
KW Clone phrAR1; DNA; retinoic acid receptor; ligand
KW complexes; human.
XX
OS Homo sapiens (Human).
XX
FH Key Location/Qualifiers
FT CDS 103..149
FT /*tag= a
XX
XX W08905355-A.
XX
PD 15-JUN-1989.
XX
PF 01-DEC-1988; 88WO-US04284.
XX
XX 02-DEC-1987; 87US-0276536.
```

```
XX (SALK ) SALK INST FOR BIOLOGICAL STUD.
XX PA
XX Evans RM, Giguere V, Ong ES, Segui PS;
XX WPI: 1989-192701/26.
XX DR P-PSDB; AAP90395.
XX
XX DNA encoding retinoic acid receptor proteins
PT - used to produce proteins for studying complexes with
PT ligands and in diagnostic assays.
XX PS
XX Disclosure; fig. 1B; 75pp; English.
XX
XX DNA of clone phrAR1 contg. the primary sequence of a
CC protein (see corresp. AAP90395) that has ligand binding and
CC transcription activating properties of retinoic acid receptor
CC (RAR) protein. Used to make chimeric receptors,
CC to produce receptor, to study binding complexes, and to screen cpds.
CC for RAR-agonists and antagonists.
XX
XX Sequence 2940 BP; 594 A; 1010 C; 790 G; 546 T; 0 other;
SQ
Query Match 17.4%; Score 174; DB 10; Length 2940;
Best Local Similarity 100.0%; Pred. No. 7.1e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGACAACGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTCACTGA 887
DB 732 GAACAACAGCTCAGACAACGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTCACTGA 791
QY 888 ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGCTTCAC 947
DB 792 ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGCTTCAC 851
QY 948 CACCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGGGACAT 1001
DB 852 CACCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGGGACAT 905
```

Search completed: March 30, 2003, 10:14:03
Job time : 214.299 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 09:12:13 ; Search time 33.7584 Seconds
(without alignments)
9093.534 Million cell updates/sec

Title: US-09-691-220-3_COPY_10000_11000
Perfect score: 1001
Sequence: 1 cccaggagactgcagctgg.....aaggctgctgcctggacat 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	174	17.4	704	2	US-08-592-383-5
2	174	17.4	2658	2	US-08-592-383-3
3	174	17.4	2328	2	US-08-095-728B-3
4	174	17.4	2928	5	PCT-US92-02320A-3
5	174	17.4	2940	2	US-08-592-383-1
6	174	17.4	2940	6	5171671-1
7	174	17.4	3036	1	US-08-306-691B-52
8	174	17.4	3036	2	US-08-095-728B-1
9	174	17.4	3036	5	PCT-US92-02320A-1
10	174	17.4	3511	3	US-08-892-747-13
11	112.8	11.3	2989	6	5223606-1
12	109.2	10.9	1576	6	5260432-1
13	86.6	8.7	558	2	US-08-896-365-5
14	56.1	5.6	1934	4	US-08-776-844-1
15	56	5.6	1959	1	US-08-342-411A-3
16	56	5.6	1959	5	PCT-US94-12883-4
17	55	5.5	7218	1	US-08-232-463-14
18	53.8	5.4	1893	6	5438126-1
19	53	5.3	320	4	US-09-165-264-7
20	52.8	5.3	1860	2	US-08-372-652-7
21	52.8	5.3	1860	5	PCT-US95-16311-7
22	52	5.2	320	4	US-09-165-264-14
23	51.8	5.2	320	4	US-09-165-264-13
24	51.2	5.1	318	4	US-09-165-264-12
25	50.4	5.0	320	4	US-09-165-264-11
26	49.6	5.0	319	4	US-09-165-264-8
27	49.6	5.0	1688	2	US-08-649-619B-2

28	49.6	5.0	1813	5	PCT-US94-12883-3	Sequence 3, Appl
29	49.6	5.0	1898	1	US-08-342-411A-1	Sequence 1, Appl
30	49.6	5.0	1979	2	US-08-649-619B-1	Sequence 1, Appl
31	49.6	5.0	2030	1	US-08-330-518-1	Sequence 1, Appl
32	49.6	5.0	2030	1	US-08-330-283-1	Sequence 1, Appl
33	49.6	5.0	2030	2	US-08-646-248-1	Sequence 1, Appl
34	49.6	5.0	2030	5	PCT-US95-13924-1	Sequence 1, Appl
35	49.6	5.0	2030	5	PCT-US95-13931-1	Sequence 1, Appl
36	47	4.7	152331	3	US-09-128-155-16	Sequence 16, Appl
37	45.8	4.6	816	1	US-08-485-971-22	Sequence 22, Appl
38	45.8	4.6	816	1	US-08-383-754-22	Sequence 22, Appl
39	45.8	4.6	816	1	US-08-485-978-22	Sequence 22, Appl
40	45.8	4.6	816	2	US-08-486-814-22	Sequence 22, Appl
41	45.8	4.6	816	3	US-08-487-472-22	Sequence 22, Appl
42	45.8	4.6	816	3	US-08-485-740-22	Sequence 22, Appl
43	45.8	4.6	816	3	US-09-162-184-22	Sequence 22, Appl
44	45.8	4.6	816	4	US-09-161-902-22	Sequence 22, Appl
45	45.8	4.6	816	4	US-09-489-777A-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-08-592-383-5
; Sequence 5, Application US/08592383
; Patent No. 5830760
; GENERAL INFORMATION:
; APPLICANT: Tsai, S. and S.J. Collins
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acti
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage
; COMPUTER: IBM PC/386 Compatible
; OPERATING SYSTEM: MS-DOS 4.01
; SOFTWARE: Word for Windows 5.01-t
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,383
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/099,242
; FILING DATE: July 28, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Broderick, Thomas F.
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: PHCR-1-7190
; TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 704 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: page 11; RAR-alpha RA binding region; positions 692 to 1395
; IMMEDIATE SOURCE:
; LIBRARY: cDNA
; US-08-592-383-5

Query Match 17.4%; Score 174; DB 2; Length 704;
Best Local Similarity 100.0%; Pred. No. 4.3e-34;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAGTGA 887

Db 41 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 100
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGGCCCAAGCAGCTGCCCGGCTTCAC 947
Db 101 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGGCCCAAGCAGCTGCCCGGCTTCAC 160
QY 948 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGGCTCCCTGCTGGGACAT 1001
Db 161 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGGCTCCCTGCTGGGACAT 214

RESULT 2

US-08-592-383-3
; Sequence 3, Application US/08592383
; Patent No. 5830760
; GENERAL INFORMATION:
; APPLICANT: Tsai, S. and S.J. Collins
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Recept
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage
; COMPUTER: IBM PC/386 Compatible
; OPERATING SYSTEM: MS-DOS 4.01
; SOFTWARE: Word for Windows 5.01-t
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,383
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/099,242
; FILING DATE: July 28, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Broderick, Thomas F.
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: FHCR-1-7190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2658 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: page 4, RAR-alpha403 dominant negative; deleted of 1311-1596 of SEQ
; IMMEDIATE SOURCE:
; LIBRARY: cDNA
US-08-592-383-3

Query Match 17.4%; Score 174; DB 2; Length 2658;
Best Local Similarity 100.0%; Pred. No. 6.6e-34;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887
Db 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 791
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGGCCCAAGCAGCTGCCCGGCTTCAC 947
Db 792 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGGCCCAAGCAGCTGCCCGGCTTCAC 851
QY 948 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGGCTCCCTGCTGGGACAT 1001
Db 852 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGGCTCCCTGCTGGGACAT 905

RESULT 3

US-08-095-728B-3
; Sequence 3, Application US/08095728B
; Patent No. 5843642
; GENERAL INFORMATION:
; APPLICANT: DMITROVSKY, ETHAN
; APPLICANT: WARRELL JR, RAYMOND P
; APPLICANT: MILLER JR, WILSON H
; APPLICANT: FRANKEL, STANLEY
; TITLE OF INVENTION: METHODS FOR THE DETECTION AND
; TITLE OF INVENTION: TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM LLP
; STREET: 1185 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,728B
; FILING DATE: 21-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/673,838
; FILING DATE: 22-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WHITE, JOHN P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 38694-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: hRAR ALPHA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 103..1488
; OTHER INFORMATION:
US-08-095-728B-3

Query Match 17.4%; Score 174; DB 2; Length 2928;
Best Local Similarity 100.0%; Pred. No. 6.8e-34;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887
Db 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 791
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGGCTTCAC 947
Db 792 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGGCTTCAC 851
QY 948 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGGCTCCCTGCTGGGACAT 1001
Db 852 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGGCTCCCTGCTGGGACAT 905

RESULT 4
PCT-US92-02320A-3

Sequence 3, Application PC/TUS9202320A
GENERAL INFORMATION:
APPLICANT: Sloan-Kettering Institute, For Cancer Research
TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOPER & DUNHAM
STREET: 30 ROCKEFELLER PLAZA
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/02320A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 675,084
FILING DATE: 22-MAR-1991
PRIOR APPLICATION DATA: US 673,838
FILING DATE: 22-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WHITE, JOHN P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 38694-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 644-0525
TELEX: (212) 422523 COOP UI
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2928 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: hRAR ALPHA
FEATURE:
NAME/KEY: CDS
LOCATION: 103..1488
OTHER INFORMATION:
PCT-US92-02320A-3
Query Match 17.4%; Score 174; DB 5; Length 2928;
Best Local Similarity 100.0%; Pred. No. 6.8e-34;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGAACAGCTCTCTCTGGACATTGACCTCTGGGACAGTTTCAGTGA 887
Db 732 GAACAACAGCTCAGAACAGCTCTCTCTGGACATTGACCTCTGGGACAGTTTCAGTGA 791
QY 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 947
Db 792 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 851
QY 948 CACCCTCACCATCGCGGACCATACCCCTCTCAAGGCTGCCTGCCTGGACAT 1001
Db 852 CACCCTCACCATCGCGGACCATACCCCTCTCAAGGCTGCCTGCCTGGACAT 905
RESULT 5
US-08-592-383-1
Sequence 1, Application US/08592383
Patent No. 5830760
GENERAL INFORMATION:
APPLICANT: Tsai, S. and S.J. Collins

TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Ac
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Centre, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage
COMPUTER: IBM PC/386 Compatible
OPERATING SYSTEM: MS-DOS 4.01
SOFTWARE: Word for Windows 5.01-t
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,383
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/099,242
FILING DATE: July 28, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Broderick, Thomas F.
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: FPCR-1-7190
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2940 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: see Figure 16B; RAR-alpha
IMMEDIATE SOURCE:
LIBRARY: cDNA
US-08-592-383-1
Query Match 17.4%; Score 174; DB 2; Length 2940;
Best Local Similarity 100.0%; Pred. No. 6.8e-34;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGAACAGCTGTCTCTCTGGACATTGACCTCTGGGACAGTTTCAGTGA 887
Db 732 GAACAACAGCTCAGAACAGCTGTCTCTCTGGACATTGACCTCTGGGACAGTTTCAGTGA 791
QY 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 947
Db 792 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 851
QY 948 CACCCTCACCATCGCGGACCATACCCCTCTCAAGGCTGCCTGCCTGGACAT 1001
Db 852 CACCCTCACCATCGCGGACCATACCCCTCTCAAGGCTGCCTGCCTGGACAT 905
RESULT 6
5171671-1
Patent No. 5171671
APPLICANT: EVANS, RONALD M.; ONG, ESTELITA S.; SEGUI,
PRUDIMAR S.; THOMPSON, CATHERINE C.; UEMSONO, KAZUHIKO
GUGUERE, VINCENT
TITLE OF INVENTION: RETINOIC ACID RECEPTOR COMPOSITION
NUMBER OF SEQUENCES: 2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/546,256
FILING DATE: 06-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 276,536
FILING DATE: 30-NOV-1988
APPLICATION NUMBER: 128,331
FILING DATE: 02-DEC-1987
SEQ ID NO:1;

LENGTH: 2940
5171671-1

Query Match 17.4%; Score 174; DB 6; Length 2940;
Best Local Similarity 100.0%; Pred. No. 6.8e-34;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA 887
|||||
Db 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA 791
|||||
QY 888 ACTCTCCACCAAGTCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947
|||||
Db 792 ACTCTCCACCAAGTCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 851
|||||
QY 948 CACCTCTACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1001
|||||
Db 852 CACCTCTACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 905
|||||

RESULT 7

US-08-306-691B-52
; Sequence 52, Application US/08306691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavorigna & Monaco, P.C.
; STREET: Two Penn Center, Suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/306,691B
FILING DATE: September 15, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
TELEX: No. 5734039e
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 3036 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-306-691B-52
Query Match 17.4%; Score 174; DB 1; Length 3036;
Best Local Similarity 100.0%; Pred. No. 6.9e-34;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA 887
|||||
Db 1701 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA 1760
|||||

QY 888 ACTCTCCACCAAGTCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947
|||||
Db 1761 ACTCTCCACCAAGTCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 1820
|||||
QY 948 CACCTCTACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1001
|||||
Db 1821 CACCTCTACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1874
|||||

RESULT 8

US-08-095-728B-1
; Sequence 1, Application US/08095728B
; Patent No. 5843642
; GENERAL INFORMATION:
; APPLICANT: DMITROVSKY, ETHAN
; APPLICANT: WARRELL JR, RAYMOND P
; APPLICANT: MILLER JR, WILSON H
; APPLICANT: FRANKEL, STANLEY
; TITLE OF INVENTION: METHODS FOR THE DETECTION AND
; TITLE OF INVENTION: TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM LLP
; STREET: 1185 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,728B
FILING DATE: 21-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/673,838
FILING DATE: 22-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WHITE, JOHN P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 38694-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3036 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: MYL-RAR
FEATURE:
NAME/KEY: CDS
LOCATION: 67..2457
OTHER INFORMATION:
US-08-095-728B-1

Query Match 17.4%; Score 174; DB 2; Length 3036;
Best Local Similarity 100.0%; Pred. No. 6.9e-34;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA 887
|||||
Db 1701 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA 1760
|||||
QY 888 ACTCTCCACCAAGTCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947
|||||
Db 1761 ACTCTCCACCAAGTCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 1820
|||||

QY 948 CACCCTCACCATCGCGGACAGATCACCCTCTCAAGGCTGCCTGGACAT 1001
|||||
Db 1821 CACCCTCACCATCGCGGACAGATCACCCTCTCAAGGCTGCCTGGACAT 1874
|||||

RESULT 9
PCT-US92-02320A-1
; Sequence 1, Application PC/TUS9202320A
; GENERAL INFORMATION:
; APPLICANT: Sloan-Kettering Institute, For Cancer Research
; TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM
; STREET: 30 ROCKEFELLER PLAZA
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10112

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/02320A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 675,084
; FILING DATE: 22-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 673,838
; FILING DATE: 22-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WHITE, JOHN P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 38694-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 644-0525
; TELEX: (212) 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3036 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: MYL-RAR
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 67..2457
; OTHER INFORMATION:
; PCT-US92-02320A-1

Query Match 17.4%; Score 174; DB 5; Length 3036;
Best Local Similarity 100.0%; Pred. No. 6.9e-34;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAGTTCAAGTTCAGTGA 887
|||||
Db 1701 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAGTTCAAGTTCAGTGA 1760
|||||
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCCGCAACAGCTGCCCGGCTTCAC 947
|||||
Db 1761 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCCGCAACAGCTGCCCGGCTTCAC 1820
|||||
QY 948 CACCCTCACCATCGCGGACAGATCACCCTCTCAAGGCTGCCTGGACAT 1001
|||||
Db 1821 CACCCTCACCATCGCGGACAGATCACCCTCTCAAGGCTGCCTGGACAT 1874
|||||

RESULT 10
US-08-892-747-13
; Sequence 13, Application US/08892747
; Patent No. 6057153
; GENERAL INFORMATION:
; APPLICANT: Shaji T. George, Michael Ma, Martina Werner,
; APPLICANT: Umberto Pace and Allan R. Goldberg
; TITLE OF INVENTION: Stabilized External Guide Sequences
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/892,747
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/372,556
; FILING DATE: January 13, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00513
; FILING DATE: January 19, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: ILI109CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..3511
; OTHER INFORMATION: /function= "PML-RAR' DNA Sequence."
; US-08-892-747-13

Query Match 17.4%; Score 174; DB 3; Length 3511;
Best Local Similarity 100.0%; Pred. No. 7.2e-34;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTACCTCTGGGACAGTTCAAGTTCAGTGA 887
|||||
Db 2176 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTACCTCTGGGACAGTTCAAGTTCAGTGA 2235
|||||
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCCGCAACAGCTGCCCGGCTTCAC 947
|||||
Db 2236 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCCGCAACAGCTGCCCGGCTTCAC 2295
|||||
QY 948 CACCCTCACCATCGCGGACAGATCACCCTCTCAAGGCTGCCTGGACAT 1001
|||||
Db 2296 CACCCTCACCATCGCGGACAGATCACCCTCTCAAGGCTGCCTGGACAT 2349
|||||

; APPLICANT: Messer, Lori A.
 ; APPLICANT: Tun-Ping, Yu
 ; TITLE OF INVENTION: GENES AND GENETIC MARKERS FOR IMPROVED
 ; TITLE OF INVENTION: REPRODUCTIVE TRAITS IN ANIMALS
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ;

Query Match 8.7%; Score 86.6; DB 2; Length 558;
Best Local Similarity 74.2%; Pred. No. 1.4e-12;
Matches 135; Conservative 0; Mismatches 45; Indels

RESULT 14
US-08-776-844-1
; Sequence 1, Application US/08776844
; Patent No. 6277976
; GENERAL INFORMATION:
; APPLICANT: ENMARK, EVA

RESULT 11
5223606-1
; Patent NO. 5223606
; APPLICANT: BLAUDIN DE THE, HUGHES; MARCHIO, AGNES; TIOLLAIS,

```

;SEQ ID NO:1: 1576
; LENGTH: 5260432-1.

```

US-08-896-365-5 ; Sequence 5, Application US/08896365
; Patent No. 5939264
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Max F.
; APPLICANT: Tuggle, Christopher K

APPLICANT: GUSTAFSSON, JAN
TITLE OF INVENTION: OR-1 ON ORPHAN RECEPTOR BELONGING
TO THE NUCLEAR RECEPTOR FAMILY
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff
STREET: 1001 G Street, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,844
FILING DATE: 24-JUN-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/03247
FILING DATE: 16-AUG-1995
APPLICATION NUMBER: UK 9413536.2
FILING DATE: 16-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A
REGISTRATION NUMBER: 32141
REFERENCE/DOCKET NUMBER: 00487.04029
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1934 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-776-844-1
Query Match 5.6%; Score 56; DB 4; Length 1934;
Best Local Similarity 63.2%; Pred. No. 7.7e-05;
Matches 86; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 866 ACCTCTGGGACAGTTTCAGTGAACCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGG 925
Db 942 AACGCTTTGCCCACTTCACTAGCTAGCTAGCTATCTCAGTCCAGGAGATCGTGACTTCG 1001
QY 926 CCAAGCAGGTGCGCGGTTTCCACCACTCCACCATCGCCGACCATCACCCTCTCAAGG 985
Db 1002 CCAAGCAGGTGCGCGGTTTCTTCAGCTGGCGGGAGGACCATGATCGCCCTCTGAAGG 1061
QY 986 CTGCGCTGCTGGACAT 1001
Db 1062 CATCCACCATCGAGAT 1077

RESULT 15
US-08-342-411A-3
Sequence 3, Application US/08342411A
Patent No. 5639616
GENERAL INFORMATION:
APPLICANT: LIAO, Shutsung
APPLICANT: SONG, Ching
TITLE OF INVENTION: UBIQUITOUS NUCLEAR RECEPTOR:
COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston

STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/342,411A
FILING DATE: 18-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KITCHELL, BARBARA S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: ARCD154
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1959 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 256..1584
US-08-342-411A-3
Query Match 5.6%; Score 56; DB 1; Length 1959;
Best Local Similarity 63.2%; Pred. No. 7.8e-05;
Matches 86; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 866 ACCTCTGGGACAGTTTCAGTGAACCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGG 925
Db 998 AACGCTTTGCCCACTTCACTAGCTAGCTAGCTATCTCAGTCCAGGAGATCGTGACTTCG 1057
QY 926 CCAAGCAGCTGCCCGGCTTCACCACTCCACCATCGCCGACCATCACCCTCTCAAGG 985
Db 1058 CCAAGCAGCTGCCCGGCTTCCTGCAGCTGGCGGGAGGACCATGATCGCCCTCTGAAGG 1117
QY 986 CTGCGCTGCTGGACAT 1001
Db 1118 CATCCACCATCGAGAT 1133
Search completed: March 30, 2003, 13:56:34
Job time : 45.7584 secs

Title: GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 13:12:08 ; Search time 75.7689 Seconds
(without alignments)
11242.392 Million cell updates/sec

Title: US-09-691-220-3_COPY_10000_11000
Perfect score: 1001
Sequence: 1 cccaggagactgcagctgg.....aaggctgctgcctggacat 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

1: /cgn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2.6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2.6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2.6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2.6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2.6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2.6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2.6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2.6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2.6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2.6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2.6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2.6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2.6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	174	17.4	2907	10	US-09-954-456-524
2	166	16.6	393	10	US-09-960-352-4521
3	166	16.6	411	10	US-09-960-352-5714
4	166	16.6	416	10	US-09-960-352-2138
5	166	16.6	417	10	US-09-960-352-12806
6	166	16.6	429	10	US-09-960-352-31
7	117.6	11.7	1375	9	US-10-239-804-7
8	117.6	11.7	1399	9	US-10-239-804-8
9	117.6	11.7	9127	9	US-10-239-804-9
10	117.6	11.7	9151	9	US-10-239-804-10
11	109.2	10.9	1577	10	US-09-797-727-1
12	99.6	10.0	848	10	US-09-833-381-1210
13	89	8.9	456	10	US-09-833-381-92
14	56	5.6	1065	10	US-09-804-682-33
15	56	5.6	1934	10	US-09-909-446-1
16	56	5.6	1934	10	US-09-909-325-1
17	56	5.6	1934	10	US-09-909-326-1
18	53.8	5.4	2288	10	US-09-962-436-567
19	52.8	5.3	1841	12	US-10-013-823-1

20	51.8	5.2	12733	9	US-10-032-393-47
21	51.8	5.2	12739	9	US-10-032-393-8
22	49.6	5.0	794	9	US-10-043-487-53
23	49.6	5.0	1622	10	US-09-925-297-176
24	47.2	4.7	1064	10	US-09-804-682-29
25	47	4.7	152331	9	US-10-095-407-16
26	46.4	4.6	1390	10	US-09-760-364-8
27	45.8	4.6	816	9	US-09-921-650-22
28	45.8	4.6	816	10	US-09-874-389-22
29	45.8	4.6	1443	9	US-09-921-650-37
30	44.6	4.5	440	9	US-10-184-644-202
31	44.6	4.5	1332	10	US-09-883-093-1
32	44.4	4.4	987	10	US-09-804-682-20
33	44.2	4.4	428	10	US-09-864-761-23488
34	44.2	4.4	577	10	US-09-864-761-6755
35	44.2	4.4	693	12	US-10-044-090-315
36	44.2	4.4	1528	10	US-09-962-832-109
37	44.2	4.4	1528	10	US-09-880-107-3328
38	44.2	4.4	1679	12	US-10-044-090-316
39	44	4.4	197496	9	US-09-877-177-10
40	43.6	4.4	450	10	US-09-960-352-10528
41	42.2	4.2	2581	10	US-09-742-732-1
42	42	4.2	405	10	US-09-960-352-13861
43	42	4.2	1450	10	US-09-880-107-3875
44	41.6	4.2	451	10	US-09-864-761-19039
45	41.6	4.2	458	10	US-09-864-761-2304

ALIGNMENTS

RESULT 1

US-09-954-456-524
; Sequence 524, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Us:
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 524
; LENGTH: 2907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-524

Query Match 17.4%; Score 174; DB 10; Length 2907;
Best Local Similarity 100.0%; Pred. No. 1.2e-38;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGTCTCTGGACATTGACCTCTGGGCAAGTTTCAGTGA 887
|||||
DB 732 GAACAACAGCTCAGAACACGCTGTCTCTGGACATTGACCTCTGGGCAAGTTTCAGTGA 791
|||||
QY 888 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947
|||||
DB 792 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 851
|||||
QY 948 CACCTCCACCATCGCCGACAGATCACCTCTCAAGGCTGCTCGCTGGACAT 1001
|||||
DB 852 CACCTCCACCATCGCCGACAGATCACCTCTCAAGGCTGCTCGCTGGACAT 905
|||||

RESULT 2
US-09-960-352-4521
; Sequence 4521, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4521
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 20-LIB188-012-Q1-E1-E11
US-09-960-352-4521

Query Match 16.6%; Score 166; DB 10; Length 393;
Best Local Similarity 97.1%; Pred. No. 1.4e-36;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGTCTCTGGACATTGACCTCTGGGCAAGTTTCAGTGA 887
|||||
DB 206 GAACAACAGCTCAGAACACGCTGTCTCTGGACATTGACCTCTGGGCAAGTTTCAGTGA 265
|||||
QY 888 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947
|||||
DB 266 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 325
|||||
QY 948 CACCTCCACCATCGCCGACAGATCACCTCTCAAGGCTGCTCGCTGGACAT 1001
|||||
DB 326 CACCTCCACCATCGCCGACAGATCACCTCTCAAGGCTGCTCGCTGGATAT 379
|||||

RESULT 3
US-09-960-352-5714
; Sequence 5714, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5714
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 25-LIB188-001-Q1-E1-C1

US-09-960-352-5714

Query Match 16.6%; Score 166; DB 10; Length 411;
Best Local Similarity 97.1%; Pred. No. 1.4e-36;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGTCTCTGGACATTGACCTCTGGGCAAGTTTCAGTGA 887
|||||
DB 206 GAACAACAGCTCAGAACACGCTGTCTCTGGACATTGACCTCTGGGCAAGTTTCAGTGA 265
|||||
QY 888 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947
|||||
DB 266 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 325
|||||
QY 948 CACCTCCACCATCGCCGACAGATCACCTCTCAAGGCTGCTCGCTGGACAT 1001
|||||
DB 326 CACCTCCACCATCGCCGACAGATCACCTCTCAAGGCTGCTCGCTGGATAT 379
|||||

RESULT 4
US-09-960-352-2138
; Sequence 2138, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 2138
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (76)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 10-LIB188-020-Q1-E1-C5
US-09-960-352-2138

Query Match 16.6%; Score 166; DB 10; Length 416;
Best Local Similarity 97.1%; Pred. No. 1.4e-36;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGTCTCTGGACATTGACCTCTGGGCAAGTTTCAGTGA 887
|||||
DB 142 GAACAACAGCTCAGAACACGCTGTCTCTGGACATTGACCTCTGGGCAAGTTTCAGTGA 201
|||||
QY 888 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947
|||||
DB 202 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 261
|||||
QY 948 CACCTCCACCATCGCCGACAGATCACCTCTCAAGGCTGCTCGCTGGACAT 1001
|||||
DB 262 CACCTCCACCATCGCCGACAGATCACCTCTCAAGGCTGCTCGCTGGATAT 315
|||||

RESULT 5
US-09-960-352-12806/c
; Sequence 12806, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12806
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 25-LIB188-001-Q1-E1-C1

FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 12806
LENGTH: 417
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 55-LIB188-005-Q1-E1-F4
US-09-960-352-12806

Query Match : 16.6%; Score 166; DB 10; Length 417;
Best Local Similarity 97.1%; Pred. No. 1.4e-36;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 887
DB 212 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 153
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGGCAAGCAGCTGCCCGGCTTCAC 947
DB 152 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGGCAAGCAGCTGCCCGGCTTCAC 93
QY 948 CACCCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1001
DB 92 CACCCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGATAT 39

RESULT 6
US-09-960-352-31
Sequence 31, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengping
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 31
LENGTH: 429
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 01-LIB188-009-Q1-E1-A1
US-09-960-352-31

Query Match : 16.6%; Score 166; DB 10; Length 429;
Best Local Similarity 97.1%; Pred. No. 1.4e-36;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 887
DB 234 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 293
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGGCAAGCAGCTGCCCGGCTTCAC 947
DB 294 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGGCAAGCAGCTGCCCGGCTTCAC 353
QY 948 CACCCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1001
DB 354 CACCCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGATAT 407

RESULT 7
US-09-960-352-31
Sequence 31, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengping
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 31
LENGTH: 429
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 01-LIB188-009-Q1-E1-A1
US-09-960-352-31

APPLICANT: Kingsman, Alan J
APPLICANT: Maden, Malcolm
APPLICANT: Corcoran, Jonathan PT
TITLE OF INVENTION: Factor
FILE REFERENCE: P009156WOCTH
CURRENT APPLICATION NUMBER: US/10/239,804
CURRENT FILING DATE: 2002-09-23
PRIOR APPLICATION NUMBER: PCT/GB00/01211
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: GB 0024300.6
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 1375
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: RARbeta2 PCR
OTHER INFORMATION: product
US-10-239-804-7

Query Match 11.7%; Score 117.6; DB 9; Length 1375;
Best Local Similarity 78.3%; Pred. No. 3.9e-23;
Matches 141; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 822 CCCCAGAACACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTT 881
DB 619 CACCACGAATTCCAGCGCTGACACCGGCTCGGATGGACCTGGGCAAAAT 678
QY 882 CAGTGAAGTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGGCAAGCAGCTGCCCGG 941
DB 679 CAGTGAAGTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGGCAAGCAGCTGCCCGG 738
QY 942 CTTCCACCACTCACCACGCTGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1001
DB 739 CTTCCACAGGCTGACCATCGGACGACATCACCTCTCAAGGCTGCTGCTGGATAT 798

RESULT 8
US-10-239-804-8
Sequence 8, Application US/10239804
Publication No. US20030053991A1
GENERAL INFORMATION:
APPLICANT: Oxford Biomedica (UK) Limited
APPLICANT: Kingsman, Alan J
APPLICANT: Maden, Malcolm
APPLICANT: Corcoran, Jonathan PT
TITLE OF INVENTION: Factor
FILE REFERENCE: P009156WOCTH
CURRENT APPLICATION NUMBER: US/10/239,804
CURRENT FILING DATE: 2002-09-23
PRIOR APPLICATION NUMBER: PCT/GB00/01211
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: GB 0024300.6
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 1399
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: FLAG RARbeta2
OTHER INFORMATION: PCR product
US-10-239-804-8

Query Match 11.7%; Score 117.6; DB 9; Length 1399;
Best Local Similarity 78.3%; Pred. No. 4e-23;
Matches 141; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 822 CCCCAGAACACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTT 881
DB 739 CTTCCACAGGCTGACCATCGGACGACATCACCTCTCAAGGCTGCTGCTGGATAT 798

Db 643 CACCACGAATTCAGCGCTGACCACCGGTCGCGATTGGACTTGGGCTCTGGGACAAATT 702
QY 882 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 941
Db 703 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 762
QY 942 CTTACACCGCTCACCACCGCGACGACATCACCCTCCCTCAAGGCTGCCTGGCCTGGACAT 1001
Db 763 CTTACAGGCTCTGACCATCGACACGACGATCACCCTGCTCAAGACCGCTGCTGGATAT 822

RESULT 9
US-10-239-804-9
; Sequence 9, Application US/10239804
; Publication No. US20030053991A1
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; APPLICANT: Kingsman, Alan J
; APPLICANT: Maden, Malcolm
; APPLICANT: Corcoran, Jonathan PT
; TITLE OF INVENTION: Factor
; FILE REFERENCE: P009156WOCTH
; CURRENT APPLICATION NUMBER: US/10/239,804
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: PCT/GB00/01211
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: GB 0024300.6
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 9127
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PONY-RARBeta2
; OTHER INFORMATION: vector genome plasmid
US-10-239-804-9

Query Match 11.7%; Score 117.6; DB 9; Length 9127;
Best Local Similarity 78.3%; Pred. No. 5.3e-23;
Matches 141; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 822 CCCCAGAACACAGCTCAGAACAGCTGTCTCTCTGGACATTCAGCTCTGGGACAAATT 881
Db 3443 CACCACGAATTCAGCGCTGACCCGCGGTCCGATTGGAGTTCGGGCTCTGGGACAAATT 3502
QY 882 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 941
Db 3503 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 3562
QY 942 CTTACACCGCTCACCACCGCGACGACATCACCCTCCCTCAAGGCTGCCTGGCCTGGACAT 1001
Db 3563 CTTACAGGCTCTGACCATCGACACGACGACATCACCCTGCTCAAGACCGCTGCTGGATAT 3622

RESULT 10
US-10-239-804-10
; Sequence 10, Application US/10239804
; Publication No. US20030053991A1
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; APPLICANT: Kingsman, Alan J
; APPLICANT: Maden, Malcolm
; APPLICANT: Corcoran, Jonathan PT
; TITLE OF INVENTION: Factor
; FILE REFERENCE: P009156WOCTH
; CURRENT APPLICATION NUMBER: US/10/239,804
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: PCT/GB00/01211
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: GB 0024300.6
; PRIOR FILING DATE: 2000-10-04

; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 9151
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: PONY-FLAG-RARBeta2 vector genome plasmid
US-10-239-804-10

Query Match 11.7%; Score 117.6; DB 9; Length 9151;
Best Local Similarity 78.3%; Pred. No. 5.3e-23;
Matches 141; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 822 CCCCAGAACACAGCTCAGAACAGCTGTCTCTCTGGACATTCAGCTCTGGGACAAATT 881
Db 3467 CACCACGAATTCAGCGCTGACCCGCGGTCCGATTGGAGTTCGGGCTCTGGGACAAATT 3526
QY 882 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 941
Db 3527 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 3586
QY 942 CTTACACCGCTCACCACCGCGACGACATCACCCTCCCTCAAGGCTGCCTGGGACAT 1001
Db 3587 CTTACAGGCTCTGACCATCGACACGACGATCACCCTGCTCAAGACCGCTGCTGGATAT 3646

RESULT 11
US-09-797-727-1
; Sequence 1, Application US/09797727
; Patent No. US20020077457A1
; GENERAL INFORMATION:
; APPLICANT: The Salk Institute for Biological Studies
; APPLICANT: TAKAKU, Fumimaro
; TITLE OF INVENTION: GAMMA RETINOIC ACID RECEPTOR
; FILE REFERENCE: SALK1150-3
; CURRENT APPLICATION NUMBER: US/09/797,727
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 08/486,325
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/100,039
; PRIOR FILING DATE: 1993-07-30
; PRIOR APPLICATION NUMBER: PCT/US90/03564
; PRIOR FILING DATE: 1990-06-22
; PRIOR APPLICATION NUMBER: US 07/370,407
; PRIOR FILING DATE: 1989-06-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1577
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Human Retinoic Acid Receptor-gamma (hRAR-gamma)
; NAME/KEY: CDS
; LOCATION: (200)..(1576)
US-09-797-727-1

Query Match 10.9%; Score 109.2; DB 10; Length 1577;
Best Local Similarity 75.8%; Pred. No. 8.4e-21;
Matches 135; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
QY 824 CCCCAGAACACAGCTCAGAACAGCTGTCTCTCTGGACATTCAGCTCTGGGACAAATTCA 883
Db 831 CCACGAAGTCCAGTGCAGACACCGCGTGCAGCTGGATCTGGGCTGTGGGACAAATTCA 890
QY 884 GTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCT 943
Db 891 GTGAGCTGGCTACCAAGTGCATCATCAAGATCGTGGAGTTCGCCAAGCGGTTGCTGCTGCT 950
QY 944 TCACACCGCTCACCACGCGGACGACGATCACCCTCTCTCAAGGCTGCCTGCCTGGACAT 1001

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:06:43 ; Search time 1290.57 Seconds
(without alignments)
12561.622 Million cell updates/sec

Title: US-09-691-220-3_COPY_10000_11000

Perfect score: 1001

Sequence: 1 cccaggagactgcagctgg.....aaggctgctgctggacat 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits: satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: gb_gss:**

18: em_gss_hum:**

19: em_gss_inv:**

20: em_gss_pln:**

21: em_gss_vrt:**

22: em_gss_fun:**

23: em_gss_mam:**

24: em_gss_mus:**

25: em_gss_other:**

26: em_gss_pro:**

27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	362.6	36.2	631	13	BI016677 MR4-ET013
2	174	17.4	560	14	BM817785 K-EST0084
3	174	17.4	646	14	BM790595 K-EST0070
4	174	17.4	885	14	BQ706025 AGENCOURT
5	174	17.4	982	12	BE794377
6	174	17.4	992	9	AL556392

c	7	172.4	17.2	700	13	BI907041
	8	170.4	17.0	173	12	BF848632
	9	170	17.0	458	14	R73335
	10	169	16.9	367	14	R71970
	11	167.6	16.7	424	12	BF721043
	12	167.6	16.7	764	9	AA790328
	13	166.6	16.6	1073	13	BM544324
	14	166.2	16.6	866	12	BF182871
	15	164.6	16.4	861	10	BE547412
	16	164.4	16.4	315	13	BM256373
	17	161.4	16.1	510	14	BQ323167
c	18	161	16.1	287	12	BF087589
	19	157.2	15.7	743	13	BI830768
	20	155.2	15.5	812	13	BI914043
	21	155.2	15.5	1034	12	BE871903
	22	154.4	15.4	330	12	BE938020
	23	152.6	15.2	426	12	BF229524
c	24	151.8	15.2	284	14	BQ329046
	25	147.2	14.7	533	10	BE650995
	26	145.8	14.6	328	12	BF333905
	27	145	14.5	670	12	BG337888
c	28	140.8	14.1	456	10	AW015525
	29	135.4	13.5	298	12	BF822393
c	30	131	13.1	270	10	BE075762
	31	129.4	12.9	302	9	AI904566
c	32	128.2	12.8	651	9	AL595869
	33	126.4	12.6	501	17	FR0007368
c	34	124.8	12.5	914	17	CNS01SPP
	35	122	12.2	841	9	AI323021
	36	121	12.1	702	13	BI827961
	37	117.6	11.7	613	9	AA543451
	38	117.6	11.7	677	12	BE854363
	39	117.6	11.7	710	12	BE854385
c	40	116.4	11.6	456	17	AQ586629
	41	114	11.4	138	10	BE483678
	42	114	11.4	347	17	BH079319
	43	112.8	11.3	513	14	BQ559158
	44	112.8	11.3	587	13	BJ008997
	45	112.8	11.3	636	17	AG050726

ALIGNMENTS

RESULT 1	BI016677	631 bp	mrna	linear	EST 13-JUN-2001
LOCUS	MR4-ET0138-190301-004-h03	ET0138	Homo sapiens	cdna	mrna sequence.
DEFINITION	BI016677				
ACCESSION	BI016677.1	GI:14420748			
VERSION	EST.				
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 631)				
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7)			3491-3496 (2000)
MEDLINE	20202663				
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001				

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR4&t2=MR4-ET0138-190301-004-h03&t3=2001-03-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 19
High quality sequence stop: 551.

FEATURES

source
1. .631
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0138"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 114 a 160 c 229 g 128 t
ORIGIN

Query Match 36.2%; Score 362.6; DB 13; Length 631;
Best Local Similarity 98.7%; Pred. No. 2.6e-72;
Matches 376; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 136 ATGCGGCTGCTATGGGTGGGGTGGTGTGTCAGGCGCCACAGCTGTGCTCATGGG 195
|||||
DB 205 ATGCGGTTGGCTATGGGTGGGGTGGTGTGTCAGGGCCACAGCTGTGCTATGGG 264
QY 196 GCTTCTGGGGCAGAACTTGATGTGGTGGTGGGATGAGGGCTGGAGTGCCTGGC 255
|||||
DB 265 GCTTCTGGGGCAGAACTTGATGTGGTGGTGGGATGAGGGCTGGAGTGCCTGGC 324
QY 256 ATGCTTGGCTCCGCGCCGTAACGCGTGTGTCGCGGTCTTACAGCTGGTGGTGCACCT 315
DB 325 AATGCTTGGCTCCGCGCGTGAACGCGTGTGTCGCGGTCTTACAGCTGGTGGTGCACCT 384
QY 316 CTTACAGCTGGCAGCTCTCTGTACAGGCTGGGGTGGGAGAGCGCTGTGAGCGCTGCA 375
DB 385 CTTACAGCTGGCAGCTCTCTGTACAGGCTGGGGTGGGAGAGCGCTGTGAGCGCTGCA 444
QY 376 GCTGCCCTTTAACCCCTCTGCCCTCCACAGCTGTGAGAAAGACCGCAAGAGAA 435
|||||
DB 445 GCTGCCCTTTAAGCCCTCTGCCCTCCACAGCTGTGAGAAAGACCGCAAGAGAA 504
QY 436 GAAGAGGTGCCCAAGCCCGAGTGTCTGTAGAGCTACAGCTGACGCCGAGGTGGGGA 495
DB 505 GAAGAGGTGCCCAAGCCCGAGTGTCTGTAGAGCTACAGCTGACGCCGAGGTGGGGA 564
QY 496 GCTCATTGAGAAGGTGCGCAA 516
DB 565 -CTCATTGAGAATGTCGCAA 584

RESULT 2
LOCUS BM817785 560 bp mRNA linear EST 06-MAR-2002
DEFINITION K-EST0084263 S21SNUS20s1 Homo sapiens cDNA clone S21SNUS20s1-2-F03
5', mRNA sequence.

ACCESSION BM817785
VERSION BM817785.1 GI:19174198
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 560)
AUTHORS Kim,N.S.; Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 2 row: F column: 03
High quality sequence stop: 560.
Location/Qualifiers
1. .560
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="S21SNUS20s1-2-F03"
/clone_lib="S21SNUS20s1"
/sex="F"
/tissue_type="Stomach"
/cell_type="Floating aggregates"
/cell_line="SNU-520"
/lab_host="Top10F"

FEATURES

source
1. .560
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="S21SNUS20s1-2-F03"
/clone_lib="S21SNUS20s1"
/sex="F"
/tissue_type="Stomach"
/cell_type="Floating aggregates"
/cell_line="SNU-520"
/lab_host="Top10F"

/note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 - 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dt)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transfection of the
remaining DNA into competent cells E. coli Top10F' with
electroporation method."

BASE COUNT .118 a 187 c 160 g .95 t
ORIGIN

Query Match 17.4%; Score 174; DB 14; Length 560;
Best Local Similarity 100.0%; Pred. No. 2.1e-29;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACAACTGCTCTCTGGACATTTGACCTCTGGGACAAAGTTTCAGTGA 887
DB 88 GAACAACAGCTCAGAACAACTGCTCTCTGGACATTTGACCTCTGGGACAAAGTTTCAGTGA 147
QY 888 ACTCTCCACCAAGTGCATCATTAAGAGCTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCCAC 947
DB 148 ACTCTCCACCAAGTGCATCATTAAGAGCTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCCAC 207
QY 948 CACCTTCACCATGCCCGACAGATCACCTCTCAAGCTGCCTGCCTGGACAT 1001
DB 208 CACCTTCACCATGCCCGACAGATCACCTCTCAAGCTGCCTGCCTGGACAT 261

RESULT 3
BM790595

```

LOCUS      BM790595                646 bp    mRNA    linear    EST 05-MAR-2002
DEFINITION K-EST0070459 S21SNU520 Homo sapiens cDNA clone S21SNU520-11-G11 5',
            mRNA sequence.
ACCESSION  BM790595
VERSION     BM790595.1  GI:19138827
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 646)
AUTHORS     Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
TITLE       21C Frontier Korean EST Project 2001
JOURNAL     Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 11 row: G column: 11
            High quality sequence stop: 646.
FEATURES   Location/Qualifiers
            1..646
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="S21SNU520-11-G11"
                /clone_lib="S21SNU520"
                /sex="F"
                /tissue_type="Stomach"
                /cell_type="Floating aggregates"
                /lab_host="Top10p"
                /note="Organ: Stomach; Vector: pTV18p1; Site_1: EcoRI;
                Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                bacterial alkaline phosphatase (BAP) and then decapped
                with tabacco acid pyrophosphatase (TAP). The decapped
                intact mRNA was ligated with DNA-RNA linker including EcoR
                I site by treatment of T4 RNA ligase and the first strand
                cDNA was synthesized from oligo dt-selected mRNA by
                priming with dt-tailed vector. The dt-tailed vector was
                adjusted to have about 60nt. The cDNA vector was
                circularized with E. coli DNA ligase after digestion of
                EcoRI which site is also included in vector. An RNA strand
                converted to a DNA strand by Okayama-Berg method. The
                obtained cDNA vectors were used for transformation of
                competent cells E. coli Top10p by electroporation method.
                The cDNA libraries constructed by this method are
                full-length enriched cDNA library."
BASE COUNT  140 a 213 c 184 g 109 t
ORIGIN
Query Match      17.4%; Score 174; DB 14; Length 646;
Best Local Similarity 100.0%; Pred. No. 2.2e-29;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  828  GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTCACCTCTGGGACAGTTTCAGTGA 887
      |||||||
Db   87  GAACAACAGCTCAGAACACACGTGTCTCTCTGGACATTCACCTCTGGGACAGTTTCAGTGA 146

QY  888  ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 947
      |||||||
Db   147  ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 206

QY  948  CACCTCACCATCGCCGACCATGACCTCTCTCAAGCTGCTCCCTGGACAT 1001
      |||||||
Db   207  CACCTCACCATCGCCGACCATGACCTCTCTCAAGCTGCTCCCTGGACAT 260

LOCUS      BE794377                982 bp    mRNA    linear    EST 20-SEP-2000
DEFINITION 601589560F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3943719 5',
            mRNA sequence.
ACCESSION  BE794377
VERSION     BE794377.1  GI:10215562
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

BQ706025                885 bp    mRNA    linear    EST 16-JUL-2002
LOCUS      AGENCOURT_8351580 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6282394
DEFINITION 5', mRNA sequence.
ACCESSION  BQ706025
VERSION     BQ706025.1  GI:21844924
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 885)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: Dr. Mark Watson
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LCM2476 row: 1 column: 11
            High quality sequence stop: 599.
FEATURES   Location/Qualifiers
            1..885
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:6282394"
                /clone_lib="NIH_MGC_113"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GCACAG(G). Library constructed by Ling Hong in the
                laboratory of Gerald M. Rubin (University of California,
                Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                Superscript II RT (Life Technologies). Note: this is a
                NIH_MGC Library."
BASE COUNT  157 a 337 c 234 g 154 t 3 others
ORIGIN
Query Match      17.4%; Score 174; DB 14; Length 885;
Best Local Similarity 100.0%; Pred. No. 2.3e-29;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  828  GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTCACCTCTGGGACAGTTTCAGTGA 887
      |||||||
Db   64  GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTCACCTCTGGGACAGTTTCAGTGA 123

QY  888  ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 947
      |||||||
Db   124  ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 183

QY  948  CACCTCACCATCGCCGACCATGACCTCTCTCAAGCTGCTCCCTGGACAT 1001
      |||||||
Db   184  CACCTCACCATCGCCGACCATGACCTCTCTCAAGCTGCTCCCTGGACAT 237

RESULT 5
BE794377
LOCUS      601589560F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3943719 5',
DEFINITION mRNA sequence.
ACCESSION  BE794377
VERSION     BE794377.1  GI:10215562
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 982)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LCM/99 row: g column: 16
 High quality sequence stop: 860.

FEATURES

source

1. 982
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3943719"
 /clone_lib="NIH_MGC_7"
 /tissue_type="small cell carcinoma"
 /cell_line="MGC3"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 231 a 283 c 317 g 151 t
 ORIGIN
 Query Match 17.4%; Score 174; DB 12; Length 982;
 Best Local Similarity 100.0%; Pred. No. 2.3e-29;
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887
 Db 381 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 440
 QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 947
 Db 441 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 500
 QY 948 CACCTCACCATCGCGACCATCACCTCTCTCAAGGCTGCTGCTGGACAT 1001
 Db 501 CACCTCACCATCGCGACCATCACCTCTCTCAAGGCTGCTGCTGGACAT 554

RESULT 6

AL556392

LOCUS

AL556392 LTI_NFL006_PL2 992 bp mRNA linear EST 16-FEB-2001
 prime, mRNA sequence.

ACCESSION

AL556392

VERSION

AL556392.1 GI:12899025

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 992)

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. 992

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DK004YG05"
 /clone_lib="LTI_NFL006_PL2"
 /tissue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com>

BASE COUNT 228 a 319 c 269 g 173 t
 ORIGIN

Query Match 17.4%; Score 174; DB 9; Length 992;
 Best Local Similarity 100.0%; Pred. No. 2.3e-29;
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887
 Db 488 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 547
 QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 947
 Db 548 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 607
 QY 948 CACCTCACCATCGCGACCATCACCTCTCTCAAGGCTGCTGCTGGACAT 1001
 Db 608 CACCTCACCATCGCGACCATCACCTCTCTCAAGGCTGCTGCTGGACAT 661

RESULT 7

BI907041

LOCUS

BI907041 NIH_MGC_118 700 bp mRNA linear EST 16-OCT-2001
 603065004F1
 mRNA sequence.

ACCESSION

BI907041

VERSION

BI907041.1 GI:16169824

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM11537 row: h column: 01

High quality sequence stop: 623.

Location/Qualifiers

1. 700

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5214000"

/clone_lib="NIH_MGC_118"

/tissue_type="Leukocyte"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV

(destroyed); RNA source leukocytes from anonymous pool of

non-activated adult donors. Library is oligo-dT primed

FEATURES

source

and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 230. Library constructed by Bento Soares and M.Fatima Bonaldo."

```

BASE COUNT      95 a  154 c  115 g  91 t  3 others
ORIGIN

Query Match      17.0%; Score 170; DB 14; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.7e-28;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGACAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAGTTCAGTGA 887
|||||
Db 50 GAACAACAGCTCAGACAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAGTTCAGTGA 109
|||||
QY 888 ACTCTCCACCAAGTCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 947
|||||
Db 110 ACTCTCCACCAAGTCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 169
|||||
QY 948 CACCCCTCACCATCGCGACACAGATCACCCCTCCTCAAGGCTGCCTGCCTGG 997
|||||
Db 170 CACCCCTCACCATCGCGACACAGATCACCCCTCCTCAAGGCTGCCTGCCTGG 219
|||||

```

```

RESULT 10
R71970
LOCUS
DEFINITION
YJ84e07.r1 Soares breast 2NBHst Homo sapiens cDNA clone
IMAGE:155460 5' similar to gb:M73779 RETINOIC ACID RECEPTOR ALPHA-1
(HUMAN);, mRNA sequence.
R71970
ACCESSION
VERSION
KEYWORDS
SOURCE
EST.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 367)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston,
R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 1457
High quality sequence stops: 291
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1457 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 291.
Location/Qualifiers
1..367
/organism="Homo sapiens"
/db_xref="GDB:573359"
/db_xref="taxon:9606"
/clone="IMAGE:155460"
/clone_lib="Soares breast 2NBHst"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors

```

```

FEATURES
source
1..367
/organism="Homo sapiens"
/db_xref="GDB:573359"
/db_xref="taxon:9606"
/clone="IMAGE:155460"
/clone_lib="Soares breast 2NBHst"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors

```

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 230. Library constructed by Bento Soares and M.Fatima Bonaldo."

```

BASE COUNT      81 a  132 c  80 g  70 t  4 others
ORIGIN

Query Match      16.9%; Score 169; DB 14; Length 367;
Best Local Similarity 99.4%; Pred. No. 2.7e-28;
Matches 169; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGACAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAGTTCAGTGA 887
|||||
Db 50 GAACAACAGCTCAGACAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAGTTCAGTGA 109
|||||
QY 888 ACTCTCCACCAAGTCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 947
|||||
Db 110 ACTCTCCACCAAGTCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 169
|||||
QY 948 CACCCCTCACCATCGCGACACAGATCACCCCTCCTCAAGGCTGCCTGCCTGG 997
|||||
Db 170 CACCCCTCACCATCGCGACACAGATCACCCCTCCTCAAGGCTGCCTGCCTGG 219
|||||

```

```

RESULT 11
BF721043
LOCUS
DEFINITION
mab62b03.y1 Soares_thymus_2NDMT Mus musculus cDNA clone
IMAGE:3974740 5' similar to SW:RRA_MOUSE P11416 RETINOIC ACID
RECEPTOR ALPHA ;, mRNA sequence.
BF721043
ACCESSION
VERSION
KEYWORDS
SOURCE
EST.
ORGANISM
house mouse
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1. (bases 1 to 424)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapb@femail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1474772
Seq primer: -40RP from Gibco
High quality sequence stop: 361.
Location/Qualifiers
1..424
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3974740"
/clone_lib="Soares_thymus_2NDMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."
113 a  135 c  103 g  73 t

```

```

FEATURES
source
1..424
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3974740"
/clone_lib="Soares_thymus_2NDMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."
113 a  135 c  103 g  73 t

```



```
Db 391 GCCGACTGCAGAGTCTTTGAAGTGGGCATGTCCAAAGGAGTCTGTGAGAAACGACCGAA 450
Qy 426 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTCTGAGAGCTACAGCTGACGCCGG 485
Db 451 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTCTGAGAGCTACAGCTGACGCCGG 510
Qy 486 AGGTGGGGAGCTCATTTGAGAGGTGCGCAAAAGCGCACAGGAAACCTTCCCTGCCCTCT 545
Db 511 AGGTGGGGAGCTCATTTGAGAGGTGCGCAAAAGCGCACAGGAAACCTTCCCTGCCCTCT 570
Qy 546 GCCAGCTGGGCAAAATACACTACG 568
Db 571 GCCAGCTGGGCAAAATACACTACG 593

RESULT 14
LOCUS BF182871 866 bp mRNA linear EST 31-OCT-2000
DEFINITION 601809396F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4040303 5',
mRNA sequence.
ACCESSION BF182871
VERSION BF182871.1 GI:11061015
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 866)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: DCTD/DP/Gazdar
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM857 row: o column: 24
High quality sequence stop: 723.
FEATURES
Location/Qualifiers
source
1. 866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4040303"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 215 a 251 c 264 g 136 t
ORIGIN
Query Match 16.6%; Score 166.2; DB 12; Length 866;
Best Local Similarity 88.7%; Pred. No. 1.4e-27;
Matches 180; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 366 GCAGCTGCAGTGCCTTTAAACCCCTCTGCCCTCCACAGCTGTGAGAAACGACCGAA 425
Db 257 GCCGACTGCAGAGTGTGTTGAGTGGGCATGTCCAAAGGAGTCTGTGAGAAACGACCGAA 316
Qy 426 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTGAGAGCTACAGCTGACGCCGG 485
Db 317 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTGAGAGCTACAGCTGACGCCGG 376
```

```
Qy 486 AGCTGGGGAGTCAATTGAGAGGTGCGCAAAAGCGCACAGGAAACCTTCCCTGCCCTCT 545
Db 377 AGTGGGGAGTCAATTGAGAGGTGCGCAAAAGCGCACAGGAAACCTTCCCTGCCCTCT 436
Qy 546 GCCAGCTGGGCAAAATACACTACG 568
Db 437 GCCAGCTGGGCAAAATACACTACG 459

RESULT 15
LOCUS BE547412 861 bp mRNA linear EST 09-AUG-2000
DEFINITION 601072989F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3458974 5',
mRNA sequence.
ACCESSION BE547412
VERSION BE547412.1 GI:9776057
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 861)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ARCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM8451 row: a column: 23
High quality sequence stop: 571.
FEATURES
Location/Qualifiers
source
1. 861
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3458974"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 247 a 270 c 219 g 125 t
ORIGIN
Query Match 16.4%; Score 164.6; DB 10; Length 861;
Best Local Similarity 88.2%; Pred. No. 3.1e-27;
Matches 179; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 366 GCAGCTGCAGTGCCTTCTTAACCCCTCTGCCCTCCACAGCTGTGAGAAACGACCGAA 425
Db 120 GCCGACTGCAGAGTGTGTTGAGTGGGCATGTCTGAGAGGAGTCTGTGAGAAACGACCGAA 179
Qy 426 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTGAGAGCTACAGCTGACGCCGG 485
Db 180 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTGAGAGCTACAGCTGACGCCGG 239
Qy 486 AGTGGGGAGTCAATTGAGAGGTGCGCAAAAGCGCACAGGAAACCTTCCCTGCCCTCT 545
Db 240 AGTGGGGAGTCAATTGAGAGGTGCGCAAAAGCGCACAGGAAACCTTCCCTGCCCTCT 299
Qy 546 GCCAGCTGGGCAAAATACACTACG 568
Db 300 GCCAGCTGGGCAAAATACACTACG 322

Search completed: March 30, 2003, 13:48:26
```

Job time : 1300.57 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:00:38 ; Search time 2122.53 seconds
(without alignments)
13725.086 Million cell updates/sec

Title: US-09-691-220-3_COPY_15000_16000
Perfect score: 1001
Sequence: 1 gctataaattcccggttt.....ccctccctccactggagaa 1001

Scoring table: IDENTITY_NUC
; Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.nam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	871.8	87.1	2907	6	AX333042	AX333042 Sequence
2	871.8	87.1	2907	9	HSRRA	X06614 Human mRNA
3	871.8	87.1	2908	6	I08117	I08117 Sequence 1
4	871.8	87.1	2928	6	AR061122	AR061122 Sequence
5	871.8	87.1	2940	6	AR052136	AR052136 Sequence
c	871.8	87.1	138999	9	AC090426	AC090426 Homo sapi
c	871.8	87.1	173441	2	AC080112	AC080112 Homo sapi
8	870.2	86.9	2940	6	I09348	I09348 Sequence 1
9	868.4	86.8	166368	2	AC018629	AC018629 Homo sapi
c	858.2	85.7	158766	2	AC015851	AC015851 Homo sapi
c	851.4	85.1	188574	2	AC131063	AC131063 Homo sapi
c	851.2	85.0	190309	2	AC126392	AC126392 Homo sapi
c	840	83.9	187557	2	AC069032	AC069032 Homo sapi
14	689	68.8	2658	6	AR052137	AR052137 Sequence
15	682.8	68.2	1081	9	HSRRA1A8	AF088895 Homo sapi
16	648.2	64.8	3036	6	AR061121	AR061121 Sequence
17	648.2	64.8	3036	6	I96215	I96215 Sequence 52
18	648.2	64.8	3036	9	HUMPMLRAR	M73779 Human PML-R
19	647.2	64.7	2008	9	AK098172	AK098172 Homo sapi
20	645.4	64.5	2300	9	HSB05408	AL834159 Homo sapi
21	639.4	63.9	2441	9	BC008727	BC008727 Homo sapi
22	373	37.3	1920	9	HSRAR	X06538 Human mRNA
23	373	37.3	1944	9	HSU41743	U41743 Human nucle
24	373	37.3	2073	9	HSU41742	U41742 Human nucle
25	273.4	27.3	231248	2	AL591067	AL591067 Mus muscu
26	263.2	26.3	2748	10	AY046943	AY046943 Mesocrice
27	258.6	25.8	162393	2	AC111746	AC111746 Rattus no
28	203.4	20.3	2130	10	RNU15211	U15211 Rattus norv
29	184	18.4	808	11	GI0521	GI0521 human STS C
30	112.8	11.3	1284	9	S50916	S50916 PML-RAR fus
31	102.2	10.2	2061	10	S56656	S56656 retinoic ac
32	79.8	8.0	1572	12	AF242868	AF242868 Synthetic
33	79.8	8.0	2085	12	AF242867	AF242867 Synthetic
34	79	7.9	1474	10	MSRARA	Y10094 M.spretus m
35	76	7.6	93	6	AR052141	AR052141 Sequence
c	74.6	7.5	76141	2	AC024393	AC024393 Homo sapi
37	65.6	6.6	179608	2	AC128497	AC128497 Rattus no
38	63.4	6.3	125020	9	AF429315	AF429315 Homo sapi
39	63.2	6.3	194776	2	AC129706	AC129706 Rattus no
40	62.6	6.3	180109	2	AC125916	AC125916 Rattus no
41	61.4	6.1	171777	2	AC073151	AC073151 Mus muscu
42	61	6.1	194929	2	AC119084	AC119084 Rattus no
43	60.2	6.0	192926	2	AC116254	AC116254 Rattus no
44	60	6.0	1494	5	GGRAA21	X78335 G.gallus re
45	60	6.0	1552	5	GGRAA1	X73972 G.gallus RA

ALIGNMENTS

RESULT 1	AX333042	Sequence	3551 from Patent WO0194629.	DNA	linear	PAT 09-JAN-2002
LOCUS	AX333042	Sequence	3551 from Patent WO0194629.	DNA	linear	PAT 09-JAN-2002
DEFINITION	AX333042	Sequence	3551 from Patent WO0194629.	DNA	linear	PAT 09-JAN-2002
ACCESSION	AX333042	Sequence	3551 from Patent WO0194629.	DNA	linear	PAT 09-JAN-2002
VERSION	AX333042.1	GI:18123676				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1					
AUTHORS	Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,					
	Horrigian, S., Soppet, D.R. and Weaver, Z.					
TITLE	Cancer gene determination and therapeutic screening using signature					

[illegible][illegible]

[illegible]

```
Db 67223 TTGATCAAGACACCCCTCTGCCAGCTCACACATCTTTCATCACAGCAAAAGCCAGGAC 67164
QY 557 TTGCTCCCTCCATCCTCAGAACTCACAGCCATTTCTCCAGCTGGGGAACCTCAACCT 616
Db 67163 TTGCTCCCTCCATCCTCAGAACTCACAGCCATTTCTCCAGCTGGGGAACCTCAACCT 67104
QY 617 CCCCCCTGCTCGTGGTGTGACAGAGGGGTGGGACAGGGGGGGGTCCCCCTGTAC 676
Db 67103 CCCCCCTGCTCGTGGTGTGACAGAGGGGTGGGACAGGGGGGGGTCCCCCTGTAC 67044
QY 677 ATACCTGCTGATATGAGGAGCTGATATTAATTTCTCGCTGGTGTGTTTTTAATTT 736
Db 67043 ATACCTGCTGATATGAGGAGCTGATATTAATTTCTCGCTGGTGTGTTTTTAATTT 66984
QY 737 TTTTGTGTTTGAATTTTAAAGAAATTTTCAATTTTAAGCACATTTATACGAAGAAAT 796
Db 66983 TTTTGTGTTTGAATTTTAAAGAAATTTTCAATTTTAAGCACATTTATACGAAGAAAT 66924
QY 797 TGTGCTGTGATATGAGGAGCTGATATTAATTTCTCGCTGGTGTGTTTTTAATTT 856
Db 66923 TGTGCTGTGATATGAGGAGCTGATATTAATTTCTCGCTGGTGTGTTTTTAATTT 66864
QY 857 GTGGCTCGGAAGGGGGCCCACTCTCTTTCATGCTGCTGCTGCTGCTGCTGCTGCT 916
Db 66863 GTGGCTCGGAAGGGGGCCCACTCTCTTTCATGCTGCTGCTGCTGCTGCTGCTGCT 66804
QY 917 CAGCCTTTTCTCTCAGTTTCTTCTTAAACACTGTAAGTACTAACTTTTCAAGGCTG 976
Db 66803 CAGCCTTTTCTCTCAGTTTCTTCTTAAACACTGTAAGTACTAACTTTTCAAGGCTG 66744
QY 977 CCTTCCCTCCCTCCCACTGGAGAA 1001
Db 66743 CCTTCCCTCCCTCCCACTGGAGAA 66719

RESULT 8
LOCUS I09348
DEFINITION Sequence 1 from Patent WO 8905355.
ACCESSION I09348
VERSION I09348.1 GI:587943
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2940)
AUTHORS Evans,R.M., Giguere,V., Ong,E.S., Segui,P.S., Umesono,K. and Thompson,C.C.
JOURNAL Patent: WO 8905355-A 1 15-JUN-1989;
FEATURES Location/Qualifiers
source 1..2940
BASE COUNT 594 a 1013 c 787 g 546 t
ORIGIN
Query Match 86.9%; Score 870.2; DB 6; Length 2940;
Best Local Similarity 94.3%; Pred. No. 1.5e-190;
Matches 872; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 77 CGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136
Db 1359 CTTGACACTCTGAGCGGACAGCGGGGGTGGGGGGGAGCGGGGGTGGCTGGCCCTG 1418
QY 137 CCGGCCAGGAGCTGTAGCCCAAGCTCAGCCCAAGCTTCCACAGAGCAGCCGCGCCAC 196
Db 1419 CCGGCCAGGAGCTGTAGCCCAAGCTCAGCCCAAGCTTCCACAGAGCAGCCGCGCCAC 1478
QY 197 CCACTCCCTGCTGACCCGACGCTGATGACACGCTGCTGCTGCTGCTGCTGCTGCTGCT 256
Db 1479 CCACTCCCTGCTGACCCGACGCTGATGACACGCTGCTGCTGCTGCTGCTGCTGCTGCT 1538
QY 257 CTTCTGCTTTTCTACCCAGCATGTGACCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 316
```

```
Db 1539 CTCTGCTTTTCTACGACCATGTGACCCCGCACAGCCCTGCCCCACCTGCCCTCCCGG 1598
QY 317 GCAGTACTGGGACCTTCCCTGGGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 376
Db 1599 GCAGTACTGGGACCTTCCCTGGGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1658
QY 377 AGGCTGGGCCCCCTAGTGGACTGCTGCTCCACAGCCCTGGGCTGAGCTCAGAGGCGGAG 436
Db 1659 AGGCTGGGCCCCCTAGTGGACTGCTGCTCCACAGCCCTGGGCTGAGCTCAGAGGCGGAG 1718
QY 437 GCCAGGAACTGAGTGGGCCCCCTGGTCTCGGGGTCTCAGGATGGGTCTTGGGGGCTCGTG 496
Db 1719 GCCAGGAACTGAGTGGGCCCCCTGGTCTCGGGGTCTCAGGATGGGTCTTGGGGGCTCGTG 1778
QY 497 TTCATCAAGACACCCCTCTGCCCCAGCTCACCACATCTTTCATCACCAGCAAAACCCAGGAC 556
Db 1779 TTCATCAAGACACCCCTCTGCCCCAGCTCACCACATCTTTCATCACCAGCAAAACCCAGGAC 1838
QY 557 TTGGCTCCCTCCATCTCAGAACTCACAAGCCATTTGCTCCCACTGGGGAACCTCAACCT 616
Db 1839 TTGGCTCCCTCCATCTCAGAACTCACAAGCCATTTGCTCCCACTGGGGAACCTCAACCT 1898
QY 617 CCCCCCTGCTCGTGGTGTGACAGAGGGGTGGGACAGGGGGGGGGTCCGCCCTGTAC 676
Db 1899 CCCCCCTGCTCGTGGTGTGACAGAGGGGTGGGACAGGGGGGGGGTCCGCCCTGTAC 1958
QY 677 ATACCTGCCATACCAACCCAGGTATTAATTTCTCGCTGCTGTTTTTAAATTTTAAATTT 736
Db 1959 ATACCTGCCATACCAACCCAGGTATTAATTTCTCGCTGCTGTTTTTAAATTTTAAATTT 2018
QY 737 TTTTCTTTTGAATTTTAAAGAAATTTTCAATTTTAAGCACATTTATACGAAGAAAT 796
Db 2019 TTTTCTTTTGAATTTTAAAGAAATTTTCAATTTTAAGCACATTTATACGAAGAAAT 2078
QY 797 TGTGCTGTATATGAGGGGAGCTGGATCCAGAGCTGGAGGGGTGGGTCGGGGGAGGGA 856
Db 2079 TGTGCTGTATATGAGGGGAGCTGGATCCAGAGCTGGAGGGGTGGGTCGGGGGAGGGA 2138
QY 857 GTGGCTCGGAAGGGGGCCCCCACTCTCTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 916
Db 2139 GTGGCTCGGAAGGGGGCCCCCACTCTCTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2198
QY 917 CAGCCTTTTCTCTCAGTTTCTTCTTAAACACTGTAAGTACTAACTTTTCAAGGCTG 976
Db 2199 CAGCCTTTTCTCTCAGTTTCTTCTTAAACACTGTAAGTACTAACTTTTCAAGGCTG 2258
QY 977 CTTTCCCTCCCTCCCACTGGAGAA 1001
Db 2259 CTTTCCCTCCCTCCCACTGGAGAA 2283

RESULT 9
AC018629
LOCUS AC018629
DEFINITION Homo sapiens chromosome 17 clone RP11-5809 map 17, *** SEQUENCING
IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC018629
VERSION AC018629.10 GI:22297481
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166368)
AUTHORS Birren,B., Nussbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone RP11-5809
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166368)
AUTHORS Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepey,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
```



```
Db 74567 ACCGCCACCCACATGGACAGCCCTCGCCCTCGCCCGCCGCTTTCTCTGCCCTTCT 74508
QY 269 ACCGACCATGTGACCCCGACAGCCCTCGCCCGACCTGCCCCCGGGGAGTACTGGGG 328
Db 74507 ACCGACCATGTGACCCCGACAGCCCTCGCCCGACCTGCCCCCGGGGAGTACTGGGG 74448
QY 329 ACCTTCCCTGGGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 388
Db 74447 ACCTTCCCTGGGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 74388
QY 389 TCAGTGGACTGCTGCTCCACAGCCCTGCGTACGCTGAGAGCCGAGGAGGAGGAGGAGGAGG 448
Db 74387 TCAGTGGACTGCTGCTCCACAGCCCTGCGTACGCTGAGAGCCGAGGAGGAGGAGGAGGAGG 74328
QY 449 GTGAGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 508
Db 74327 GTGAGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 74268
QY 509 CCCCCTGCCCCAGCTCACACATCTTCATCACCAGCAAGCCAGGACTTGGCTCCCCCA 568
Db 74267 CCCCCTGCCCCAGCTCACACATCTTCATCACCAGCAAGCCAGGACTTGGCTCCCCCA 74208
QY 569 TCCCTCAGAACTCACAAGCCATTTCTCCCGAGCTGGGAACTCAACCTCCCGCTGCTC 628
Db 74207 TCCCTCAGAACTCACAAGCCATTTCTCCCGAGCTGGGAACTCAACCTCCCGCTGCTC 74148
QY 629 GGTGTTGACAGAGGGGTTGGGACAGGGGGGGGGTTCCTCCCTGATACATACCTGCGCAT 688
Db 74147 GGTGTTGACAGAGGGGTTGGGACAGGGGGGGGGTTCCTCCCTGATACATACCTGCGCAT 74088
QY 689 ACCACCCAGGATTAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 748
Db 74087 ACCACCCAGGATTAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 74028
QY 749 TTTTAAATAAGAAATTTTCAATTTAAGCACTTTAAGCACTTTAAGCACTTTAAGCACTTT 808
Db 74027 TTTTAAATAAGAAATTTTCAATTTAAGCACTTTAAGCACTTTAAGCACTTTAAGCACTTT 73968
QY 809 TGGGGGAGCTGGATCCAGAGCTGGAGGGGGTGGGTCGGGGGAGGAGGAGTGGCTCGGAAG 868
Db 73967 TGGGGGAGCTGGATCCAGAGCTGGAGGGGGTGGGTCGGGGGAGGAGGAGTGGCTCGGAAG 73908
QY 869 GGGCCCCCACTCTCTTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 928
Db 73907 GGGCCCCCACTCTCTTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 73848
QY 929 CCTCAGTTTCTCTTTAAACTGTGAAGTACTGAAGTACTGAAGTACTGAAGTACTGAAGTACT 988
Db 73847 CCTCAGTTTCTCTTTAAACTGTGAAGTACTGAAGTACTGAAGTACTGAAGTACTGAAGTACT 73788
QY 989 TCCCACTGGAGAA 1001
Db 73787 TCCCACTGGAGAA 73775

RESULT 12
AC126392/c
LOCUS
DEFINITION
Homo sapiens chromosome 17 clone RP11-1029F16 map 17, WORKING DRAFT
SEQUENCE, 11 ordered pieces.
ACCESSION
AC126392
VERSION
AC126392.2 GI:22123095
KEYWORDS
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 190309)
REFERENCE
Birren,B., Nusbaum,C., Lander,E., Allen,A., Allen,N., Anderson,S.,
Camrata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Farro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mieng,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27553
Center clone name: 1029_F16
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 187521 bases at least Q40
Consensus quality: 188193 bases at least Q30
Consensus quality: 188537 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 189309; sum-of-contigs
Quality coverage: 15.5 in Q20 bases; agarose-fp
Quality coverage: 14.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
```

```
AUTHORS
Birren,B., Nusbaum,C., Lander,E., Allen,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camrata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Farro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mieng,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 190309)
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:21699357.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27553
Center clone name: 1029_F16
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 187521 bases at least Q40
Consensus quality: 188193 bases at least Q30
Consensus quality: 188537 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 189309; sum-of-contigs
Quality coverage: 15.5 in Q20 bases; agarose-fp
Quality coverage: 14.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
```

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* i 30513: contig of 30513 bp in length
* 30514 30613: gap of 100 bp
* 30614 31764: contig of 1151 bp in length
* 31765 31864: gap of 100 bp
* 31865 33693: contig of 1829 bp in length
* 33694 33793: gap of 100 bp
* 33794 33405: contig of 1612 bp in length
* 33406 35505: gap of 100 bp
* 35506 37691: contig of 2186 bp in length
* 37692 37791: gap of 100 bp
* 37792 51370: contig of 13579 bp in length
* 51371 51470: gap of 100 bp
* 51471 66014: contig of 14544 bp in length
* 66015 66114: gap of 100 bp
* 66115 108183: contig of 42069 bp in length
* 108184 108283: gap of 100 bp
* 108284 130510: contig of 22227 bp in length
* 130511 130610: gap of 100 bp
* 130611 166740: contig of 36130 bp in length
* 166741 166840: gap of 100 bp
* 166841 190309: contig of 23469 bp in length.

FEATURES

source
1. 190309
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-1029F16"
/clone_lib="RPC1-11 Human Male BAC"
1. 30513
/note="assembly_fragment"
clone_end:SP6
vector_side:left
30614..31764
/note="assembly_fragment"
31865..33693
/note="assembly_fragment"
33794..35405
/note="assembly_fragment"
35506..37691
/note="assembly_fragment"
37792..51370
/note="assembly_fragment"
51471..66014
/note="assembly_fragment"
66115..108183
/note="assembly_fragment"
108284..130510
/note="assembly_fragment"
130611..166740
/note="assembly_fragment"
166841..190309
/note="assembly_fragment"
clone_end:T7
vector_side:right
BASE COUNT 48207 a 48484 c 47543 g 45018 t 1057 others
ORIGIN

Query Match 85.0%; Score 851.2; DB 2; Length 190309;
Best Local Similarity 98.7%; Pred. No. 7.9e-186;
Matches 856; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 135 CCCCCGAGCAGCTGTAGCCAGCTCAGCCCCAGCTCAACAGAGAGAGCCGGCC 194
|||||
Db 130510 CCCCCCAGGGGNNAACTACCCAGCTCAGCCCCAGCTCAACAGAGAGAGCCGGCC 130451
QY 195 ACCCACTCCCGTACCCGAGCCACATGGACACAGCCCTCGCCCTCCGCGCGGCTT 254
|||||
Db 130450 ACCCACTCCCGTACCCGAGCCACATGGACACAGCCCTCGCCCTCCGCGCGGCTT 130991

QY 255 TTCTGCTCTTCTACCGACCATGTGACCCCGACACAGCCCTGCCCCACCTGCCCCCTCC 314
|||||
Db 130390 TTCTGCTCTTCTACCGACCATGTGACCCCGACACAGCCCTGCCCCACCTGCCCCCTCC 130331
QY 315 GGGCAGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGCAGCAGCTCTCTTGAC 374
Db 130330 GGGCAGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGCAGCAGCTCTCTTGAC 130271
QY 375 AGAGGCTGGGCGCTCAGTGGACTGCTGCTCCACAGCCTGGGCTGACGTGACAGGCG 434
Db 130270 AGAGGCTGGGCGCTCAGTGGACTGCTGCTCCACAGCCTGGGCTGACGTGACAGGCG 130211
QY 435 AGGCCAGGAACACTGAGTGGAGCCCTGGTCTGGGTCTCAGATGGGTCTGGGGGCGCTCG 494
Db 130210 AGGCCAGGAACACTGAGTGGAGCCCTGGTCTGGGTCTCAGATGGGTCTGGGGGCGCTCG 130151
QY 495 TGTTTCATCAAGACACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAGCCGAGG 554
Db 130150 TGTTTCATCAAGACACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAGCCGAGG 130091
QY 555 ACTTGGCTCCCCATCCTCAGAACTCACAAGCCATTTCTCCCCAGCTGGGGAACCTCAAC 614
Db 130090 ACTTGGCTCCCCATCCTCAGAACTCACAAGCCATTTCTCCCCAGCTGGGGAACCTCAAC 130031
QY 615 CTCCCCCCTGCTCGGTGGTGACAGAGGGGTGGGAGCGGGGGGTTCCTCCCTGT 674
Db 130030 CTCCCCCCTGCTCGGTGGTGACAGAGGGGTGGGAGCGGGGGGTTCCTCCCTGT 129971
QY 675 ACATACCTGCTCCATACCAACCCAGGTATTAAATTCGCTGGTCTTTTATTTAAT 734
Db 129970 ACATACCTGCTCCATACCAACCCAGGTATTAAATTCGCTGGTCTTTTATTTAAT 129911
QY 735 TTTTGTGTTTGAATTTTAAAGAAATTTTCAATTTAAGCACATTTATCTGAAGGAA 794
Db 129910 TTTTGTGTTTGAATTTTAAAGAAATTTTCAATTTAAGCACATTTATCTGAAGGAA 129851
QY 795 TTTGCTGCTGTATTGGGGAGCTGGATCCAGAGCTGGAGGGGTGGGTCGGGGGAGG 854
Db 129850 TTTGCTGCTGTATTGGGGAGCTGGATCCAGAGCTGGAGGGGTGGGTCGGGGGAGG 129791
QY 855 GAGTGGCTGGAGGGGCCCCCACTCTCCTTTTCATGTCCTGCCCCCAGTCTCTCTC 914
Db 129790 GAGTGGCTGGAGGGGCCCCCACTCTCCTTTTCATGTCCTGCCCCCAGTCTCTCTC 129731
QY 915 CTCAGCCTTTTCTCTCAGTCTTTTCTTTTAAAGTCTGAAAGTACTAACTTTTCAAGGCC 974
Db 129730 CTCAGCCTTTTCTCTCAGTCTTTTCTTTTAAAGTCTGAAAGTACTAACTTTTCAAGGCC 129671
QY 975 TGCCCTTCCCTCCCTCCCACTGGAGAA 1001
Db 129670 TGCCCTTCCCTCCCTCCCACTGGAGAA 129644

RESULT 13

AC069032/c
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-185D6 map 18, WORKING DRAFT
SEQUENCE, 44 unordered pieces.
ACCESSION
AC069032
VERSION
AC069032.4 GI:12963008
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 187557)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Homo sapiens chromosome 18, clone RP11-185D6
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 187557)
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczky, J., Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Pollara, V., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (17-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 187557)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T., Lehoczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 18, 2001 this sequence version replaced gi:11225387.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: Li0383
Center clone name: 185_D_6

----- Summary Statistics

Sequencing vector: M13; M7815; 2% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 166116 bases at least Q40
Consensus quality: 176954 bases at least Q30
Consensus quality: 180848 bases at least Q20
Insert size: 167000; agarose-
Insert size: 183257; sum-of-
Quality coverage: 3.7 in Q20 bases; agarose-
Quality coverage: 3.4 in Q20 ba.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1294: contig of 1294 bp in length
* 1295 1394: gap of 100 bp
* 1395 2505: contig of 1111 bp in length
* 2506 2605: gap of 100 bp
* 2606 3992: contig of 1387 bp in length
* 3993 4092: gap of 100 bp
* 4093 5527: contig of 1435 bp in length
* 5528 5627: gap of 100 bp
* 5628 6675: contig of 1048 bp in length
* 6676 7775: gap of 100 bp
* 7776 8026: contig of 1251 bp in length
* 8027 8126: gap of 100 bp
* 8127 9162: contig of 1036 bp in length
* 9163 9262: gap of 100 bp
* 9263 10351: contig of 1089 bp in length
* 10352 10451: gap of 100 bp
* 10452 11757: contig of 1306 bp in length
* 11758 11857: gap of 100 bp
* 11858 12878: contig of 1021 bp in length
* 12879 12978: gap of 100 bp
* 12979 14159: contig of 1181 bp in length
* 14160 14259: gap of 100 bp
* 14260 15818: contig of 1559 bp in length
* 15819 15918: gap of 100 bp
* 15919 17145: contig of 1227 bp in length
* 17146 17245: gap of 100 bp
* 17246 18765: contig of 1520 bp in length
* 18766 18865: gap of 100 bp
* 18866 20138: contig of 1273 bp in length
* 20139 20238: gap of 100 bp
* 20239 21879: contig of 1641 bp in length
* 21880 21979: gap of 100 bp
* 21980 23877: contig of 1898 bp in length
* 23878 23977: gap of 100 bp
* 23978 25680: contig of 1703 bp in length
* 25681 25780: gap of 100 bp
* 25781 28099: contig of 2319 bp in length
* 28100 28199: gap of 100 bp
* 28200 30797: contig of 2598 bp in length
* 30798 30897: gap of 100 bp
* 30898 33191: contig of 2294 bp in length
* 33192 33291: gap of 100 bp
* 33292 36520: contig of 3229 bp in length
* 36521 36620: gap of 100 bp
* 36621 38842: contig of 2222 bp in length
* 38843 38942: gap of 100 bp
* 38943 42470: contig of 3528 bp in length
* 42471 42570: gap of 100 bp
* 42571 45581: contig of 3011 bp in length
* 45582 45681: gap of 100 bp
* 45682 48005: contig of 2324 bp in length
* 48006 48105: gap of 100 bp
* 48106 51907: contig of 3802 bp in length
* 51908 52007: gap of 100 bp
* 52008 55311: contig of 3304 bp in length
* 55312 55411: gap of 100 bp
* 55412 59353: contig of 3942 bp in length
* 59354 59453: gap of 100 bp
* 59454 63506: contig of 4053 bp in length
* 63507 63606: gap of 100 bp
* 63607 68052: contig of 4446 bp in length
* 68053 68152: gap of 100 bp
* 68153 72726: contig of 4574 bp in length
* 72727 72826: gap of 100 bp
* 72827 78593: contig of 5767 bp in length
* 78594 78693: gap of 100 bp
* 78694 84391: contig of 5698 bp in length
* 84392 84491: gap of 100 bp

Db 1066 TGTCTGTGTATTGG 1081

Search completed: March 30, 2003, 13:12:03
Job time : 2833.53 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 07:57:38 ; Search time 196.299 Seconds
(without alignments)
11483.757 Million cell updates/sec

Title: US-09-691-220-3_COPY_15000_16000

Perfect score: 1001

Sequence: 1 gcctaaattcccggtt.....ccctccctccactggagaa 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	951	95.0	20512	24	Genomic DNA encodi
2	871.8	87.1	2907	24	Human cDNA differe
3	871.8	87.1	2907	24	Lung cancer relate
4	871.8	87.1	2928	20	Human RAR-alpha CD
5	871.8	87.1	2940	10	DNA of clone phRAR
6	871.8	87.1	2940	16	RAR-alpha clone la
7	870.2	86.9	2928	13	AAQ29338
8	842.8	84.2	2930	23	AAQ29308
9	689	68.8	2658	16	AAQ81477

10	648.8	64.8	2086	24	AAQ38338
11	648.2	64.8	3036	13	AAQ29334
12	648.2	64.8	3036	19	AAV20474
13	648.2	64.8	3036	20	AAV64990
14	648.2	64.8	3511	21	AAQ38656
15	645	64.4	3511	17	AAT33259
16	644.6	64.4	3511	17	AAT33246
17	578	57.7	595	24	AAQ62827
18	565.4	56.5	3052	23	AAQ83047
19	373	37.3	1944	23	AAQ83045
20	373	37.3	1944	23	AAQ83045
21	373	37.3	2073	23	AAQ85135
22	373	37.3	2457	23	AAQ83049
23	274.4	27.4	1481	17	AAT33247
24	274.4	27.4	1481	17	AAT33260
25	79.8	8.0	1692	23	AAQ83808
26	59	5.9	296	24	ABN16897
27	54.8	5.5	1000	21	AAQ02484
28	52.4	5.2	320	21	AAQ38183
29	52.4	5.2	840	24	ABQ35494
30	52.4	5.2	840	24	ABQ35495
31	52	5.2	2188	20	AAZ77506
32	51.6	5.2	1337	20	AAZ17263
33	51.2	5.1	4475	14	AAQ45995
34	51.2	5.1	10732	21	AAQ10594
35	51	5.1	600	24	ABQ52496
36	51	5.1	600	24	ABQ52497
37	50.6	5.1	401	22	AAI92118
38	50.6	5.1	690	22	AAH33877
39	50.2	5.0	1327	24	ABQ68452
40	49.6	5.0	320	21	AAQ38185
41	49.4	4.9	320	21	AAQ38186
42	49.4	4.9	387	23	ABV57769
43	49.4	4.9	434	22	AAI84343
44	49	4.9	318	21	AAQ38184
45	49	4.9	1232	22	AAQ11124

ALIGNMENTS

RESULT 1
AAL38339
ID AAL38339 standard; DNA; 20512 BP.

XX AAL38339;

XX 15-AUG-2002 (first entry)

Genomic DNA encoding the human nuclear hormone receptor protein.
Human nuclear hormone receptor; therapeutic agent; gene therapy;
immune response; chromosome 17; single nucleotide polymorphism;
gene; OS.

XX Homo sapiens.

Key Location/Qualifiers

CDS 1684..15208
/tag= a
/product= "Human nuclear hormone receptor protein"
1684..1846
/tag= b
/number= 1
1847..6808
/tag= c
/number= 1
6809..6957
/tag= d
/number= 2
6958..8277
/tag= e
/number= 2

CDNA encoding huma
myl/RAR-alpha fusi
Human PML/RAR-alpha
Human myl/RAR-alpha
Human PML/RAR-alpha
PML-RAR-alpha RNA
PML-RAR-alpha RNA
CDNA sequence #614
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
RAR-alpha RNA sequ
RAR-alpha DNA sequ
DNA encoding novel
Human ORFX polynuc
Human colon cancer
Primer used in the
Oligonucleotide fo
Oligonucleotide fo
Human ovarian tumo
Human gene express
IDUA - exons III t
Gene encoding a su
Oligonucleotide fo
Oligonucleotide fo
Human polynucleoti
Human colon cancer
Listeria monocytog
Primer used in the
Primer used in the
Human prostate exp
Human polynucleoti
Primer used in the
Human small cell l

FT exon 8278..8419
FT /tag= f
FT /number= 3
FT intron 8420..10406
FT /tag= g
FT /number= 3
FT exon 10407..10567
FT /tag= h
FT /number= 4
FT intron 10568..10827
FT /tag= i
FT /number= 4
FT exon 10828..11004
FT /tag= j
FT /number= 5
FT intron 11005..12798
FT /tag= k
FT /number= 5
FT exon 12799..13003
FT /tag= l
FT /number= 6
FT intron 13004..13759
FT /tag= m
FT /number= 6
FT exon 13760..13918
FT /tag= n
FT /number= 7
FT intron 13919..14505
FT /tag= o
FT /number= 7
FT exon 14506..14658
FT /tag= p
FT /number= 8
FT intron 14659..15143
FT /tag= q
FT /number= 8
FT exon 15144..15208
FT /tag= r
FT /number= 9
FT allele replace(4084,C)
FT /tag= s
FT allele replace(6482,A)
FT /tag= t
FT allele replace(8066,G)
FT /tag= u
FT allele replace(8699,C)
FT /tag= v
FT allele replace(12897,T)
FT /tag= w
FT allele replace(14442,C)
FT /tag= x
XX WO200231146-A2.
PN 18-APR-2002.
PD 05-OCT-2001; 2001WO-US31095.
PF 11-OCT-2000; 2000US-239117P.
PR 19-OCT-2000; 2000US-0691220.
XX (PEKE) PE CORP NY.
PA
XX Wei M, Ye J, Yan C, Ketchum KA, Di Francesco V, Beasley EM;
PI WPI; 2002-426282/45.
XX P-PSDB; AAO21489.
DR
XX New human nuclear hormone receptor proteins and nucleic acids, useful
PT as models or targets for developing human therapeutic targets, and in
PT identifying therapeutic proteins and modulators of nuclear hormone
PT receptor expression
XX

PS Claim 1; Fig 3; 73pp; English.
XX The invention relates to an isolated peptide of a novel human nuclear
CC hormone receptor with a fully defined sequence of 457 amino acids given
CC in the specification. The novel human nuclear hormone receptor peptides
CC and nucleic acids encoding them can be used as models for the development
CC of human therapeutic targets, aid in the identification of therapeutic
CC proteins, and serve as targets for the development of human therapeutic
CC agents that modulate nuclear hormone receptor activity in cells and
CC tissues that express the nuclear hormone receptor. The nucleic acids may
CC be used as a query sequence to perform searches against sequence
CC databases to identify family members or related sequences, as probes or
CC primers, to construct recombinant vectors, to identify compounds that
CC modulate nuclear hormone receptor nucleic acid expression, in gene
CC therapy, and as antisense constructs to control nuclear hormone receptor
CC gene expression in cells, tissues or organisms. The polypeptides can be
CC used to raise antibodies or to elicit an immune response, as a reagent in
CC assays designed to determine protein levels in biological fluids, as
CC markers for tissues in which a corresponding protein is expressed, to
CC identify a binding partner/ligand to develop a system for the
CC identification of inhibitors of the binding reaction, in drug screening
CC assays, and to identify compounds that modulate protein activity. This
CC polynucleotide sequence represents the genomic DNA encoding the human
CC nuclear hormone receptor protein of the invention.
XX
SQ Sequence 20512 BP; 3591 A; 5987 C; 6281 G; 4541 T; 112 other;
Query Match 95.0%; Score 951; DB 24; Length 20512;
Best Local Similarity 100.0%; Pred. No. 2.8e-198;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCTAAATAATCCCGTGTGCCACCACTCTTGAAACCCCAAGCTTGGGTGTTAATCC 60
Db 15000 GCCTAAATAATCCCGTGTGCCACCACTCTTGAAACCCCAAGCTTGGGTGTTAATCC 15059
QY 61 CGAATTGGGGGGCCCCCGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
Db 15060 CGAATTGGGGGGCCCCCGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 15119
QY 121 NNN 180
Db 15120 NNN 15179
QY 181 GAAGCAGCCGGCCACCCACTCCCGGTGACGCCCGCCACCATGGACACAGCCCTCGCC 240
Db 15180 GAAGCAGCCGGCCACCCACTCCCGGTGACGCCCGCCACCATGGACACAGCCCTCGCC 15239
QY 241 CTCGGCCCGGGCTTTTCTCTGCTTTTACCGACCATGTGACCCCGCCACCGCCCTCGCC 300
Db 15240 CTCGGCCCGGGCTTTTCTCTGCTTTTACCGACCATGTGACCCCGCCACCGCCCTCGCC 15299
QY 301 CCACCTGCTCCCGGGCAGTACTGGGGAGCTTCCCTGGGGAGCGGGGAGGAGGCA 360
Db 15300 CCACCTGCTCCCGGGCAGTACTGGGGAGCTTCCCTGGGGAGCGGGGAGGAGGCA 15359
QY 361 GCGACTCTTGGACAGAGGCTTGGGCCCTCAGTGTGCTGCTCCACAGCTGGGCT 420
Db 15360 GCGACTCTTGGACAGAGGCTTGGGCCCTCAGTGTGCTGCTCCACAGCTGGGCT 15419
QY 421 GAGCTCAGAGCGCGAGGCCAGGAACCTGAGTGAGGCCCTCTGGTCTCTCAGGATGGG 480
Db 15420 GAGCTCAGAGCGCGAGGCCAGGAACCTGAGTGAGGCCCTCTGGTCTCTCAGGATGGG 15479
QY 481 TCCTGGGGGCTCGTGTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCAC 540
Db 15480 TCCTGGGGGCTCGTGTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCAC 15539
QY 541 CAGCAACGCCAGACTTGGCTCCGCCATCTCTAGAACTCACAAGCCATTCCTCCCGCAGC 600
Db 15540 CAGCAACGCCAGACTTGGCTCCGCCATCTCTAGAACTCACAAGCCATTCCTCCCGCAGC 15599
QY 601 TGGGGACCTCAACCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 15600 TGGGGACCTCAACCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15659


```
QY 257 CTCTGCTTTCTACCGACATGTGACCGCGACCGACCGCTGCCCGCACCTGCCCTCCCG 316
Db 1539 CTCTGCTTTCTACCGACATGTGACCGCGACCGACCGCTGCCCGCACCTGCCCTCCCG 1598
QY 317 GCAGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376
Db 1599 GCAGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658
QY 377 AGGCTGGGGCCCTCAGTGGAGTGCCTGTCTCCACAGCCCTGGGCTGACGTCAGAGGCCGAG 436
Db 1659 AGGCTGGGGCCCTCAGTGGAGTGCCTGTCTCCACAGCCCTGGGCTGACGTCAGAGGCCGAG 1718
QY 437 GCCAGGAAGTGAAGTGGGCGGCTGTCTGGGTCTCAGAGTGGGCTGGGGGCGCTCGTG 496
Db 1719 GCCAGGAAGTGAAGTGGGCGGCTGTCTGGGTCTCAGAGTGGGCTGGGGGCGCTCGTG 1778
QY 497 TTCATCAAGACACCCCTGTGCCAGCTCACCACATCTTCCATCAGGAGGAGGAGGAGGAGGAG 556
Db 1779 TTCATCAAGACACCCCTGTGCCAGCTCACCACATCTTCCATCAGGAGGAGGAGGAGGAGGAG 1838
QY 557 TTGGCTCCCCCATCTCAGAACTCAGAGCCATTGCTCCCCAGCTGGGGAACCTCAACCT 616
Db 1839 TTGGCTCCCCCATCTCAGAACTCAGAGCCATTGCTCCCCAGCTGGGGAACCTCAACCT 1898
QY 617 CCCCCCTGCCGTGGTGTGACAGAGGGGTGGGACAGGGGCGGGGGTTCCTCCCTGTATC 676
Db 1899 CCCCCCTGCCGTGGTGTGACAGAGGGGTGGGACAGGGGCGGGGGTTCCTCCCTGTATC 1958
QY 677 ATACCTGCCATACCAACCCAGGATTAATCTCGCTGGTGTCTGTTTATTTATTTATTT 736
Db 1959 ATACCTGCCATACCAACCCAGGATTAATCTCGCTGGTGTCTGTTTATTTATTTATTT 2018
QY 737 TTTGTTTTGATTTTTTAAGAATTTTCAATTTAAGCACATTTATATGAAGGAAT 796
Db 2019 TTTGTTTTGATTTTTTAAGAATTTTCAATTTAAGCACATTTATATGAAGGAAT 2078
QY 797 TGTGCTGTGATTTGGGGGAGCTGATCAGAGCTGGAGGGGGTGGGTCGGGGGAGGGA 856
Db 2079 TGTGCTGTGATTTGGGGGAGCTGATCAGAGCTGGAGGGGGTGGGTCGGGGGAGGGA 2138
QY 857 GTGCTCGAAGGGGCCCCACTCCTTCATGTCCCTGTGCGCCCGCCAGTCTCCTCCT 916
Db 2139 GTGCTCGAAGGGGCCCCACTCCTTCATGTCCCTGTGCGCCCGCCAGTCTCCTCCT 2198
QY 917 CAGCCTTTTCTCCTCAGTCTTCTTTAAACTGTGAAGTACTAACTTTTCAAGGCCG 976
Db 2199 CAGCCTTTTCTCCTCAGTCTTCTTTAAACTGTGAAGTACTAACTTTTCAAGGCCG 2258
QY 977 CTTTCCCTCCCTCCCTCCACTGGAGAA 1001
Db 2259 CTTTCCCTCCCTCCCTCCACTGGAGAA 2283
```

RESULT 4

AAV64991
ID AAV64991 standard; cDNA; 2928 BP.

AAV64991;

05-FEB-1999 (first entry)

Human RAR-alpha cDNA.

Fusion protein; myl; retinoic acid receptor-alpha; RAR: human;
acute promyelocytic leukaemia; APL; t(15;17); translocation;
treatment; all-trans retinoic acid; ss.

Homo sapiens.

Key Location/Qualifiers
CDS 103..1491

/*tag= a
/product= "RAR-alpha"

/note= "retinoic acid receptor-alpha"

US5843642-A.
01-DEC-1998.
21-JUL-1993; 93US-0095728.
21-JUL-1993; 93US-0095728.
22-MAR-1991; 91US-0673838.
22-MAR-1991; 91US-0675084.

(SLOK) SLOAN KEYTERING INST CANCER RES.

Dmitrovsky E, Frankel S, Miller WH, Warrell RP;

WPI; 1999-044563/04.

P-PSDB; AAW81964.

Diagnosis of acute promyelocytic leukaemia - by detecting nucleic
acid encoding abnormal retinoic acid receptor-alpha

Disclosure; Column 29-34; 38pp; English.

This sequence encodes the human retinoic acid receptor alpha, RAR-alpha
which is used in a method for identifying a subject with acute
promyelocytic leukaemia (APL) resulting from a t(15;17) translocation who
will respond to treatment with all-trans retinoic acid. The protein can
also be used to identify a subject with indications of APL who will not
respond to treatment with all-trans retinoic acid.

Sequence 2928 BP; 582 A; 1011 C; 789 G; 546 T; 0 other;

Query Match 87.1%; Score 871.8; DB 20; Length 2928;
Best Local Similarity 94.4%; Pred. No. 3.2e-181;
Matches 873; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 77 CCGTNN 136

Db 1359 CCGTGGACATCTCAGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGGCTGGCCCC 1418

QY 137 CCGGCCAGGAGTGTAGCCCGAGCTCAGCCCGAGCTCCACAGAGAGCAGCGCGGCAC 196

Db 1419 CCGGCCAGGAGTGTAGCCCGAGCTCAGCCCGAGCTCCACAGAGAGCAGCGCGGCAC 1478

QY 197 CCACTCCCGTACCGCCCGCCAGCCACATGGACACAGCCCTCGCCCTCCGCCCGGCTTTT 256

Db 1479 CCACTCCCGTACCGCCCGCCAGCCACATGGACACAGCCCTCGCCCTCCGCCCGGCTTTT 1538

QY 257 CTCTGCTTTTCTACCGACCATGTGACCCCGCACAGCCCTGCCCCCACCTGCCCTCCCGG 316

Db 1539 CTCTGCTTTTCTACCGACCATGTGACCCCGCACAGCCCTGCCCCCACCTGCCCTCCCGG 1598

QY 317 GCAGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376

Db 1599 GCAGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658

QY 377 AGGCTGGGCGCTCAGTGGAGTGCCTGCCACAGCGCTGGGTGAGTGCAGAGGCGGAG 436

Db 1659 AGGCTGGGCGCTCAGTGGAGTGCCTGCCACAGCGCTGGGTGAGTGCAGAGGCGGAG 1718

QY 437 GCCAGGAAGTGAAGTGGGCGGCTGGTCTGGGTCTCAGAGTGGGTCTGGGGGCGCTCGTG 496

Db 1719 GCCAGGAAGTGAAGTGGGCGGCTGGTCTGGGTCTCAGAGTGGGTCTGGGGGCGCTCGTG 1778

QY 497 TTCATCAAGACACCCCTCTGCCAGCTCAGGACATCTTATCATCAGCAAGAGGAGGAGGAG 556

Db 1779 TTCATCAAGACACCCCTCTGCCAGCTCAGGACATCTTATCATCAGCAAGAGGAGGAGGAG 1838

QY 557 TTGGCTCCCCCATCTCAGAACTCAGAGCCATTGCTCCCCAGCTGGGGGAGGAGGAGGAGGAG 616

Db 1839 TTGGCTCCCCCATCTCAGAACTCAGAGCCATTGCTCCCCAGCTGGGGGAGGAGGAGGAGGAG 1898

[illegible]

Db 1728 GCCAGGAAGTGAAGAGCCCTGGTCTGGGTCTCAGGATGGTCTGGGGGCTCGTG 1787
QY 497 TTATCAAGACACCCCTCTGCCAGCTCACACACATCTTCATCACGACAAACGCCAGGAC 556
Db 1788 TTATCAAGACACCCCTCTGCCAGCTCACACACATCTTCATCACGACAAACGCCAGGAC 1847
QY 557 TTGGCTCCCCCAGCTCAGAACTCACAAGCATTGCTCCCGAGCTGGGGAACCTCAACT 616
Db 1848 TTGGCTCCCCCAGCTCAGAACTCACAAGCATTGCTCCCGAGCTGGGGAACCTCAACT 1907
QY 617 CCCCCCTGCTCGGTGGTGACAGAGGGGTGGGACAGGGGGTCCCGCTGTAC 676
Db 1908 CCCCCCTGCTCGGTGGTGACAGAGGGGTGGGACAGGGGGTTCGCCCTGTAC 1967
QY 677 ATACCTCGCCATACCAACCCAGCTATTAATCTCGTGGTGTGTTTTTATTTAATTT 736
Db 1968 ATACCTCGCCATACCAACCCAGCTATTAATCTCGTGGTGTGTTTTTATTTAATTT 2027
QY 737 TTTGTTTTGATTTTTTAAGAATTTTCATTTTAAGCACATTTATACGAAGAA 794
Db 2028 TTTGTTTTGATTTTTTAAGAATTTTCATTTTAAGCACATTTATTTAAGCAAAAAA 2085

RESULT 11
AAQ29334.
ID AAQ29334 standard; cDNA; 3036 BP.
XX
AC AAQ29334;
XX
DT 09-MAR-1993 (first entry)
XX
DE myl/RAR-alpha fusion.
XX
KW Retinoic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;
KW APL; translocation; chromosome 17; chromosome 15; PCR; primer;
KW {t(15;17)(q21;q11-22)}; breakpoint; polymerase chain reaction; ss.
XX
OS Synthetic.

PH Key Location/Qualifiers
FT CDS 67..2457
FT /*tag= a
FT primer_bind 715..744
FT /*tag= b
FT primer_bind 976..1001
FT /*tag= c
FT primer_bind 1382..1413
FT /*tag= d
XX
XX WO9216660-A.
XX
XX 01-OCT-1992.
XX
XX 23-MAR-1992; 92WO-US02320.
XX
XX 22-MAR-1991; 91US-0673838.
XX 22-MAR-1991; 91US-0675084.
XX
XX (SLOK) SLOAN KETTERING INST CANCER.
XX
XX Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;
XX Warrell RP;
XX
XX WPI; 1992-349240/42.
XX P-PSDB; AAR27533.
XX
XX Marker for acute promyelocytic leukaemia and other neoplasias -
XX comprising nucleic acid and encoded abnormal retinoic acid
XX receptor-alpha receptor
XX
XX Disclosure; Fig 4; 84pp; English.
XX
XX

CC The sequence given shows a fusion between retinoic acid receptor
CC (RAR)-alpha and myl which is characteristic of acute promyelocytic
CC leukemia (APL). This is caused by a translocation of a portion of
CC the long arm of chromosome 17 onto the long arm of chromosome 15
CC {t(15;17)(q21;q11-22)}. The breakpoint region has been cloned and it
CC has been shown that DNA rearrangements are clustered in the region
CC of the first intron of RAR-alpha. This sequence was isolated by
CC polymerase chain reaction (PCR) using primers which correspond to
CC sequences both 5' and 3' to the breakpoint region.
XX
SQ Sequence 3036 BP; 605 A; 1018 C; 890 G; 523 T; 0 other;

Query Match 64.8%; Score 648.2; DB 13; Length 3036;
Best Local Similarity 92.5%; Pred. No. 2.9e-132; Indels 0; Gaps 0;
Matches 650; Conservative 0; Mismatches 53;

QY 77 CQTNNNGGCCCC 136
Db 2328 CTTGACACTCTGAGGGACAGCCGGGGGTGGGGGGGGGGGGTGGCTGGCCCC 2387
QY 137 CCGCCAGGAGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAAACAGAGCAGCCCGGCCAC 196
Db 2388 CCGCCAGGAGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAAACAGAGCAGCCCGGCCAC 2447
QY 197 CCACCTCCCGCTGACCGCCACGCCACATGGACACAGCCCTCGCCCTCGCCCGCTTTT 256
Db 2448 CCACCTCCCGCTGACCGCCACGCCACATGGACACAGCCCTCGCCCTCGCCCGCTTTT 2507
QY 257 CTCTGCTTTTACCGACCATGTGACCCCGCACAGCCCTCGCCCGCTCGCCCTCGCCG 316
Db 2508 CTCTGCTTTTACCGACCATGTGACCCCGCACAGCCCTCGCCCGCTCGCCCTCGCCG 2567
QY 317 GCAGTACTGGGGACCTTCCTCTGGGGACGGGGAGGAGGAGGAGGAGTCTCTTGGACAG 376
Db 2568 GCAGTACTGGGGACCTTCCTCTGGGGACGGGGAGGAGGAGGAGGAGTCTCTTGGACAG 2627
QY 377 AGGCTTGGCCCTCAGTGACCTGCTGCCACAGCCCTGGGCTGAGCTCAGAGGCCGAG 436
Db 2628 AGGCTTGGCCCTCAGTGAGCTGCTGCCACAGCCCTGGGCTGAGCTCAGAGGCCGAG 2687
QY 437 GCAGGAAGTGAAGTGGCCCTGCTGGGTCTCAGATGGTCTCTGGGGGCTCGTG 496
Db 2688 GCAGGAAGTGAAGTGGCCCTGCTGGGTCTCAGATGGTCTCTGGGGGCTCGTG 2747
QY 497 TTATCAAGACACCCCTCTGCCAGCTCACACATCTTCATCACGACAAACGCCAGGAC 556
Db 2748 TTATCAAGACACCCCTCTGCCAGCTCACACATCTTCATCACGACAAACGCCAGGAC 2807
QY 557 TTGGCTCCCCATCTCAGAACTCACAAGCATTGCTCCCGAGCTGGGGAACTCAACCT 616
Db 2808 TTGGCTCCCCATCTCAGAACTCACAAGCATTGCTCCCGAGCTGGGGAACTCAACCT 2867
QY 617 CCCCCCTGCTCGGTGGTGACAGAGGGGTGGGACAGGGGGTGGGGGGTTCGCCCTGTAC 676
Db 2868 CCCCCCTGCTCGGTGGTGACAGAGGGGTGGGACAGGGGGTGGGGGGTTCGCCCTGTAC 2927
QY 677 ATACCTCGCCATACCAACCCAGGATTAATCTCGTGGTGTGTTTTTATTTAATTT 736
Db 2928 ATACCTCGCCATACCAACCCAGGATTAATCTCGTGGTGTGTTTTTATTTAATTT 2987
QY 737 TTTGTTTTGATTTTTTAAGAATTTTCATTTTAAGCAACA 779
Db 2988 TTTGTTTTGATTTTTTAAGAATTTTCATTTTAAGCAACA 3030

RESULT 12
AAV20474
ID AAV20474 standard; DNA; 3036 BP.
XX
AC AAV20474;
XX
DT 17-JUN-1998 (first entry)
XX

[illegible]

```

02-MAY-2000.
14-JUL-1997;   97US-0892747.
13-JAN-1995;   95US-0372556.
19-JAN-1996;   96WO-USO0513.
(UUYA ) UNIV YALE.
Pace U, Goldberg AR, Werner M, George ST, Ma M;
WPI; 2000-399010/34.
Novel modified external guide sequences for eukaryotic RNase P for
cleaving hepatitis viral RNA comprises RNase P targeting sequence and
recognition sequence complementary to targeted sequence in target RNA
Example 1; Column 45-48; 76pp; English.
The invention relates to modified human RNase P external guide sequences
(EGSs). The EGS is a ribozyme which forms the catalytic domain of the
ribonucleoprotein RNase P, which is involved in the processing of tRNA.
The EGS comprises a cleavage targeting sequence and a recognition
sequence which is complementary to the target RNA sequence. The
recognition sequence comprises an A and a D recognition arm which
correspond to the aminoacyl acceptor stem and the D stem of tRNA. The A
recognition arm is located at the 3' end of the EGS and the D at the 5'.
end. The cleavage targeting sequence forms a structure corresponding to
the T stem and T loop of precursor tRNA. The modified EGSs of the
invention comprise at least one modified nucleotide or an unmodified
deoxyribonucleotide, and the 3' end of the molecule is preferably
modified. The A and D arms, and the T stem and loop can consist of
ribonucleotides with 5'-phosphate or 5'-phosphorothioate linkages;
2'-modified ribonucleotides with 5'-phosphate or 5'-phosphorothioate
linkages; or deoxyribonucleotides with 5'-phosphate or 5'-
phosphorothioate linkages. EGSs of the invention are used for cleaving
a target RNA molecule. The invention discloses EGSs which are targeted
to hepatitis B virus (HBV) RNA sequences, particularly hepatitis B
surface antigen (HbsAg) RNA, and EGSs targeted at PML/retinoic acid
receptor alpha (RAR-alpha) fusion gene mRNA which is found in the
majority of acute promyelocytic leukaemia (APL) cases. Sequences
AAA38652- AAA38655 represent active and control PML/RAR-alpha fusion gene
RNA- targeted EGSs used in an exemplification of the invention. The
target RNA (AAA38651) comprises the transcript of the junction between
the PML and the RAR-alpha gene fragments of the fusion gene. The fusion
gene is generated by a balanced reciprocal chromosomal translocation
between the long arms of chromosome 15 and 17, and the fusion protein
encoded is a dominant negative regulator of myeloid differentiation.
Sequence 3511 BP; 715 A; 1168 C; 1038 G; 589 T; 1 other;
Query Match          64.8%; Score 648.2; DB 21; Length 3511;
Best Local Similarity 92.5%; Pred. No. 3.1e-132;
Matches 650; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
Qy      77 CGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGCGCC 136
       |
Db      2803 CCTGGACACTGTAGCGGCACAGCCGGGGTGCGGGCGGACGGGTGCCCTGGCCCC 2862
               |||||
Qy      137 CCCGCCAGGCAGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAACAAGACAGCCGGCCAC 196
Db      2863 CCCGCCAGGCAGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAACAAGACAGCCGGCCAC 2922
Qy      197 CCATCCCCTGTACCGCCACGCCACATGACACAGCCCTCGCCCTCGCCCGCCGCTTTT 256
       |||||||
Db      2923 CCATCCCGCTGTACCGCCACGCCACATGACACAGCCCTCGCCCTCGCCCGCCGCTTTT 2982
Qy      257 CTCTGCCTTTCTACCGACCATTGTCACCCCGCACAGCCCTGCCCCACCTGCCCTCCCGG 316
       |||||||
Db      2983 CTCCTGCCTTTCTACCGACCATTGTCACCCCGCACAGCCCTGCCCCACCTGCCCTCCCGG 3042
Qy      317 GCAGTACTGGGGACCTTCCCTCTGGGGACGGGGAGGAGGACGACCTCTTTGGACAG 376

```


Search completed: March 30, 2003, 10:14:57
Job time : 250.299 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 09:12:13 ; Search time 33.7584 Seconds
(without alignments)
9093.534 Million cell updates/sec

Title: 'US-09-691-220-3_COPY_15000_16000

Perfect score: 1001

Sequence: 1 gctctaaattcccggtt.....ccctccctcccaactggagaa 1001

Scoring table: IDENTITY_NVC

Gapex 10.0 , Capext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

: Listing first 45 summaries

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*

6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	871.8	87.1	2928	2	US-08-095-728B-3
2	871.8	87.1	2928	5	PCT-US92-02320A-3
3	871.8	87.1	2940	2	US-08-592-383-1
4	871.8	87.1	2940	6	5171671-1
5	689	68.8	2658	2	US-08-592-383-3
6	648.2	64.8	3036	1	US-08-306-691B-52
7	648.2	64.8	3036	2	US-08-095-728B-1
8	648.2	64.8	3036	5	PCT-US92-02320A-1
9	648.2	64.8	3511	3	US-08-892-747-13
10	76	7.6	93	2	US-08-592-383-8
11	52.4	5.2	320	4	US-09-165-264-11
12	50.2	5.0	319	4	US-09-165-264-8
13	49.6	5.0	320	4	US-09-165-264-13
14	49.4	4.9	320	4	US-09-165-264-7
15	49.4	4.9	320	4	US-09-165-264-14
16	49	4.9	318	4	US-09-165-264-12
17	48.8	4.9	152331	3	US-09-128-155-16
18	48.4	4.8	7218	1	US-08-232-463-14
19	43.4	4.3	4480	3	US-09-191-171-7
20	43.4	4.3	4480	4	US-09-385-707-7
21	43.2	4.3	2269	4	US-09-394-645-1
22	43.2	4.3	2269	4	US-09-243-560B-1
23	43	4.3	198	1	US-08-330-108-16
24	43	4.3	198	5	PCT-US92-10087-16
25	43	4.3	72604	4	US-09-268-992-7
26	43	4.3	72604	4	US-09-657-474-7
27	42.8	4.3	53526	3	US-08-658-136-2

28 42.8 4.3 53577 3 US-08-658-136-1 Sequence 1, Appli
29 41.8 4.2 936 4 US-09-179-558-62 Sequence 62, Appl
30 41.6 4.2 1931 3 US-09-019-942-2 Sequence 2, Appli
31 41.6 4.2 1931 4 US-09-099-041A-1 Sequence 1, Appli
32 41.6 4.2 1931 4 US-09-245-281-1 Sequence 1, Appli
33 41.6 4.2 1931 4 US-09-470-271-2 Sequence 2, Appli
34 41.6 4.2 1931 4 US-09-207-359B-1 Sequence 1, Appli
35 41.4 4.1 19124 2 US-08-487-826B-13 Sequence 13, Appl
36 41.2 4.1 1553 4 US-09-280-116-10 Sequence 10, Appl
37 41 4.1 1817 1 US-08-473-981A-5 Sequence 5, Appli
38 41 4.1 1817 2 US-08-474-087-5 Sequence 5, Appli
39 41 4.1 3602 4 US-09-402-929-1 Sequence 1, Appli
40 41 4.1 7808 2 US-08-149-097D-22 Sequence 22, Appl
41 41 4.1 7808 3 US-08-949-386-22 Sequence 22, Appl
42 41 4.1 7808 3 US-08-450-562-22 Sequence 22, Appl
43 41 4.1 7808 4 US-08-984-709A-22 Sequence 22, Appl
44 41 4.1 7808 4 US-08-450-272-22 Sequence 22, Appl
45 40.6 4.1 1364 1 US-08-265-087-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-08-095-728B-3

; Sequence 3, Application US/08095728B

; Patent No. 5843642

; GENERAL INFORMATION:

; APPLICANT: DMITROVSKY, ETHAN

; APPLICANT: WARRELL JR, RAYMOND P

; APPLICANT: MILLER JR, WILSON H

; APPLICANT: FRANKEL, STANLEY

; TITLE OF INVENTION: METHODS FOR THE DETECTION AND

; TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: COOPER & DUNHAM LLP

; STREET: 1185 AVENUE OF THE AMERICAS

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.24

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/095.728B

; FILING DATE: 21-JUL-1993

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/673,838

; FILING DATE: 22-MAR-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: WHITE, JOHN P

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 38694-A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 278-0400

; TELEFAX: (212) 391-0525

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2928 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cdna

; IMMEDIATE SOURCE:

; CLONE: hRAR ALPHA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 103..1488


```

Db 2079  TGTGCTGTGTATTGGGGGAGCTGGGATCCAGAGCTGGAGGGGGTGGGTCCGGGGGAGGGGA 2138
QY 857  GTGGCTCGGAAGGGGGCCCCCACTCTCCTTTTCATGTCCCTGTGCCCCCCCCAGTTCCTCCCT 916
Db 2139  GTGGCTCGGAAGGGGGCCCCCACTCTCCTTTTCATGTCCCTGTGCCCCCCCCAGTTCCTCCCT 2198
QY 917  CAGCCTTTTCCCTCCTCAGTTTTCTCTTTAAACTGTGAAGTACTAACTTTCCAAGGCCTG 976
Db 2199  CAGCCTTTTCCCTCCTCAGTTTTCTCTTTAAACTGTGAAGTACTAACTTTCCAAGGCCTG 2258
QY 977  CCTTCCCTCCCTCCCACTGGAGAA 1001
Db 2259  CCTTCCCTCCCTCCCACTGGAGAA 2283

RESULT 4
5171671-1
; Patent No. 5171671
; APPLICANT:: EVANS, RONALD M.; ONG, ESTELITA S.; SEGUI,
; PRUDMAR S.; THOMPSON, CATHERINE C.; UEMSONO, KAZUHIKO
; GUGUERE, VINCENT
; TITLE OF INVENTION: RETINOIC ACID RECEPTOR COMPOSITION
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/546,256
; FILING DATE: 06-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 276,536
; FILING DATE: 30-NOV-1988
; APPLICATION NUMBER: 128,331
; FILING DATE: 02-DEC-1987
; SEQ ID NO: 1:
; LENGTH: 2940
; 5171671-1

```

RESULT 5
 US-08-592-383-3
 ; Sequence 3, Application US/08592383
 ; Patent No. 5830760
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsai, S. and S. J. Collins
 ; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
 ; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98101
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage
 ; COMPUTER: IBM PC/386 Compatible
 ; OPERATING SYSTEM: MS-DOS 4.01
 ; SOFTWARE: Word for Windows 5.01-t
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/592,383
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/099,242
 ; FILING DATE: July 28, 1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Broderick, Thomas F.
 ; REGISTRATION NUMBER: 31,332
 ; REFERENCE/DOCKET NUMBER: FHC-1-7190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2658 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; DESCRIPTION: page 4, RAR-alpha403 dominant negative; deleted of 1311-1596
 ; IMMEDIATE SOURCE:

[illegible]

```

RESULT 9
US-08-892-747-13
; Sequence 13, Application us/08892747
; Patent No. 6057153
GENERAL INFORMATION:
APPLICANT: Shaji T. George, Michael Ma, Martina Werner,
APPLICANT: Umberto Pace and Allan R. Goldberg
TITLE OF INVENTION: Stabilized External Guide Sequences
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/892,747
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: 08/372,556
FILING DATE: January 13, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: ILI109CIP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 351 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..3511
OTHER INFORMATION: /function= "PML-RAR' DNA Sequence."
US-08-892-747-13

Query Match          64.8%;   Score 648.2; DB 3; Length
Best Local Similarity 92.5%;   Pred. No. 2.2e-152;
Matches 650; Conservative 0; Mismatches 55; Indels

QY    77  CGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
      | |
Db     2803 CCTGACACTCTGAGCGGCACACGCCGGGGGTGGGGGGGGGACGGCGGGGTGGTG
QY    137 CCGCCAGGAGCTGTAGCCCCAGCCCTCAGCCCAGCTCCACAGCACGACGACG
      |||||
Db     2863 CCGCCAGGAGCTGTAGCCCCAGCCCTCAGCCCCAGCTCCAACAGAAGCA
      |||||
QY    197 CCACTCCCGGTGACGGCCCGACGCCACATGGACACAGAGCCTTCGCCCTTCGCC
      |||||
Db     2923 CCACTCCCGGTGACGGCCCGACGCCACATGGACACAGACCCTTCGCCCTTCGCC
      |||||
QY    257 CTCTGCCTTTCTACCGACCATGTGTGACCCCGCACCAGCCCTGCGCCCCACCTGT

```



```

; FILE REFERENCE: 4477
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer sequence
US-09-165-264-8

```

[illegible]

```

RESULT 13
US-09-165-264-13/c
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ.ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13

```

	Query Match	5.0%;	Score 49.6;	DB 4;	Length 320;
	Best Local Similarity	42.6%;	Pred. No. 0.0017;		
	Matches 103; Conservative	0;	Mismatches 139;	Indels 0;	Gaps 0;
OY	73	CCCCCGTNNNGG	132		
Ddb	313	CCTGTGCAGTGCCCCCCCCCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC	254		
OY	133	CCCCC CGC GAG CAG CT GT AG CC CC CAG C CTC A GC CC CAG CTC CA C A G A G A G C G C C G	192		
Ddb	253	CCCCCCCCCCC GCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC	194		
OY	193	CCACCCACTCCCGGTGACC GG CCGCACGCC ACATGGACACAGCCCTTCGCCCTCCGCCCGCGGC	252		
Ddb	193	CC	134		

```

QY      253 TTTTCTCTGCCTTTCTACCGACCATGTGACCCCCCACCAGCCCTGCGCCCCACACTGCCTC 312
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      133 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 14
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      313 CC 314
          ||
Db      73 CC 72

RESULT 14
US-09-165-264-7/c
; Sequence 7, Application US/09165264
; Patent No.. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamcoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Locl Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

```

Query Match 4.9%; Score 49.4; DB 4; Length 320;
Best Local Similarity 42.3%; Pred. No. 0.0019;
Matches 107; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

[illegible]

```

RESULT 15
US-09-165-264-14/c
; Sequence 14, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiyayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer sequence

```

US-09-165-264-14

```

Query Match.      4.9%; Score 49.4; DB 4; Length 320;
Best Local Similarity 42.0%; Pred. No. 0.0019;
Matches 113; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 46 CTTGGGTGTTAATCCGAATGGGGGCCCCCGTNNNNNNNNNNNNNNNNNNNN 105
   || || || || || || || || || || || || || || || || || ||
Db 319 CTCTGCTACATGCACACCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 260
   || || || || || || || || || || || || || || || || || ||
QY 106 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 165
   || || || || || || || || || || || || || || || || || ||
Db 259 CCCC GCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 200
   || || || || || || || || || || || || || || || || || ||
QY 166 GCCCCAGCTCAACAGACAGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 225
   || || || || || || || || || || || || || || || || || ||
Db 199 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 140
   || || || || || || || || || || || || || || || || || ||
QY 226 GACACAGCCCTCGGCCCTCGGCCCTTTTCTGTGCTTTTCTACCGACCATGTGACCCC 285
   || || || || || || || || || || || || || || || || || ||
Db 139 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 80
   || || || || || || || || || || || || || || || || || ||
QY 286 GCACACGCCCTGCCGCCACCTGCCCTCCC 314
   || || || || || || || || || || || || || || || || || ||
Db 79 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 51

```

Search completed: March 30, 2003, 13:57:00
 Job time : 59.7584 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 13:12:08 ; Search time 75.7689 Seconds
(without alignments)
11242.392 Million cell updates/sec

Title: US-09-691-220-3_COPY_15000_16000
Perfect score: 1001
Sequence: 1 gcctaaattcccggtt.....ccctccctccactggagaa 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:

- 1: /cgn2_6/ptodata/1/pubnpa/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubnpa/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubnpa/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubnpa/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubnpa/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubnpa/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubnpa/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubnpa/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubnpa/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubnpa/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubnpa/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubnpa/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubnpa/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	871.8	87.1	2907	10	US-09-954-456-524
C 2	578	57.7	595	10	US-09-822-830A-614
C 3	54.4	5.4	671	9	US-10-184-644-346
C 4	52.2	5.2	12733	9	US-10-032-393-47
C 5	52.2	5.2	12739	9	US-10-032-393-8
C 6	48.8	4.9	152331	9	US-10-035-407-16
C 7	48.2	4.8	934	9	US-10-078-090-83
C 8	46	4.6	569	9	US-10-012-542-44
C 9	45.4	4.5	1064	10	US-09-804-682-29
C 10	44.4	4.4	266	10	US-09-925-301-392
C 11	44.4	4.4	1333	9	US-10-102-806-28
C 12	44	4.4	227	10	US-09-960-352-14543
C 13	44	4.4	397	10	US-09-960-352-13784
C 14	44	4.4	689	9	US-10-001-887-14
C 15	44	4.4	1545	10	US-09-960-352-14268
C 16	43.6	4.4	328	10	US-09-960-352-14268
C 17	43.4	4.3	7428	9	US-10-147-026-11
C 18	43.2	4.3	267	10	US-09-823-101-9
C 19	43.2	4.3	1014	9	US-10-102-806-305

C 20	43.2	4.3	2259	10	US-09-860-298-1	Sequence 1, Appli
C 21	43	4.3	272	10	US-09-960-352-6986	Sequence 6986, Ap
C 22	43	4.3	393	10	US-09-960-352-4582	Sequence 4582, Ap
C 23	43	4.3	748	10	US-09-910-943-361	Sequence 361, Appl
C 24	43	4.3	1468	10	US-09-764-870-40	Sequence 40, Appl
C 25	43	4.3	1998	10	US-09-764-903-19	Sequence 19, Appl
C 26	42.8	4.3	1085	10	US-09-804-682-33	Sequence 33, Appl
C 27	42.8	4.3	2756	10	US-09-925-301-351	Sequence 351, Appl
C 28	42.8	4.3	53522	9	US-09-904-968A-1	Sequence 1, Appli
C 29	42.4	4.2	694	10	US-09-846-258-3	Sequence 3, Appli
C 30	42.2	4.2	259	10	US-09-960-352-3453	Sequence 3453, Ap
C 31	42.2	4.2	259	10	US-09-960-352-7233	Sequence 7233, Ap
C 32	42.2	4.2	869	10	US-09-925-301-359	Sequence 359, App
C 33	42.2	4.2	3372	10	US-09-888-615-18	Sequence 18, Appl
C 34	42	4.2	1923	9	US-10-042-141-12	Sequence 12, Appl
C 35	42	4.2	1923	10	US-09-726-643-12	Sequence 12, Appl
C 36	42	4.2	43058	10	US-09-954-456-292	Sequence 292, App
C 37	42	4.2	43058	10	US-09-954-456-529	Sequence 529, App
C 38	42	4.2	43058	10	US-09-880-107-3950	Sequence 3950, Ap
C 39	41.8	4.2	347	10	US-09-960-352-3371	Sequence 3371, Ap
C 40	41.8	4.2	373	10	US-09-960-352-3655	Sequence 3655, Ap
C 41	41.8	4.2	413	10	US-09-960-352-2919	Sequence 2919, Ap
C 42	41.8	4.2	487	10	US-09-960-352-7373	Sequence 7373, Ap
C 43	41.8	4.2	713	9	US-10-091-483-136	Sequence 136, App
C 44	41.8	4.2	713	10	US-09-764-846-136	Sequence 136, App
C 45	41.8	4.2	936	10	US-09-974-300-836	Sequence 836, App

ALIGNMENTS

RESULT 1
US-09-954-456-524
; Sequence 524, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Usi
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 524
; LENGTH: 2907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-524

Query Match 87.1%; Score 871.8; DB 10; Length 2907;
Best Local Similarity 94.4%; Pred. No. 2.9e-211;
Matches 873; Conservative 0; Mismatches 52; Indels 0; Gaps 0;


```

: APPLICANT: Desnoyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3430R1C227
: CURRENT APPLICATION NUMBER: US/10/184,644
: CURRENT FILING DATE: 2002-06-28
: Prior Application removed - See file Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 346
: LENGTH: 871
: TYPE: PRT
: ORGANISM: Homo Sapien
: US-10-184-644-346

```

Query Match 5.4%; Score 54.4; DB 9; Length 671;
Best Local Similarity 6.9%; Pred. NO. 0.00033;
Matches 37; Conservative 182; Mismatches 316; Indels 0; Gaps 0;

Qy	23	CACCACTGCTGAACCCCAAGCTGTGGTGTATCCCGAATTGGGGGCCCCCGGTNN	82
Db	563	NNMM.BT..M.C.BYSM.YARB.TTM.KBNWTHRY.Y.MM.ATBHAM...SAB..AC	504
Qy	83	NN	142
Db	503	...TN.GYMBH.SHB.M.T.M..SD.M..M..BS..MM.MBYHKS.C.SYWAY.BWB.	444
Qy	143	AGGCAGCTGTAGCCCAAGCCTCAGCCCACTCCAACAAGAAGCACGCCGGCCACCCTC	202
Db	443	MT..M..YBYM.M.C.MYT..ASSS..MNTSMNT.Y..T...SHSSS..CYCYMCYC	384
Qy	203	CCCTGACGCCCAAGCAACATGACACAGACCCCTGCCTCCGCCCGGGCTTTCTCTGC	262
Db	383	MMB..HH.Y.HCSSCCSCY.T.CYHTV.YYY.M..M..Y.YY....YYYYT...YY	324
Qy	263	CTTTCTACCGACCATGTGACCCCAAGCAGCCCTGCCCCACCTGCCCTCCCGGGCAGTA	322
Db	323	YT..HYIYW.S.YHB.HSSHSSSSS..Y..M.MYC.Y.M..M.T.MYC..MMSBHSHS	264
Qy	323	CTGGGACCTTCCTCGGGGAGGGAGGAGGAGGAGGAGGACTCCTTGACAGAGGCCT	382
Db	263	SSSSSSSSSYKYTB...MCHSHTWHSRSTS.TMMMYCC..CY.YTYBTMM..A.	204
Qy	383	GGGCCCTCAGTGACTGCCTGCCCAAGCCTGGGCTGACGTACAGAGCGCGGAGCCAGG	442
Db	203	H.HSAM.S.SSS.SN...S.SBST.H.HSFTWTYMSBSKM.T.AMYM.CSNHSSMHSHS	144
Qy	443	AAC TGAGTGAGGCCCTGTCCTGGGTCTCAGGATGGCTCTCGGGGCCCTCGTGTTTCATC	502
Db	143	HS.KYHSHTTABTABTKBCYH.HH.HTHSCHT.NT..T.SHSSSB...TSRSTD.NN..	84
Qy	503	AAGACACCCCCTCGCCCACTCACCACATCTTCATCACCAGAAACCCAGGACT	557
Db	83	.W.C.N.CMYTN.MCRNHMGMR...HM.C.TA.DAC.....RMN..MBTCHT	29

RESULT 4

```

US-10-032-393-47/c
; Sequence 47, Application US/10032393
; Publication NO. US20030027286A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Wall, Daniel
; APPLICANT: Gross, Molly
; TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
; FILE REFERENCE: ELITRA.010A

```

TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
FILE REFERENCE: ELITRA.010A

Query Match 5.2%; Score 52.2; DB 9; Length 12739;
Best Local Similarity 42.9%; Pred. No. 0.0057;

```

; CURRENT APPLICATION NUMBER: US/10/032,393
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/259,434
; PRIOR FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 09/948,993
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 12733
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; -OTHER INFORMATION: Vector pEPEF14
US-10-032-393-47

```

Query Match 5.2%; Score 52.2; DB 9; Length 12733;
Best Local Similarity 42.9%; Pred. No. 0.0057;
Matches 111; Conservative 0; Mismatches 148; Indels 0; Cons. 0.

Qy	72	CCCCCGCTNNNGG	131
Db	4743	CCCCCGCC	4684
Qy	132	GGCCCGCCGGCAGGAGCTGATGCCAGCCTCAGCCCTCAGTCCACAGAAGCAGGCCG	191
Db	4683	CCCCCGCC	4624
Qy	192	GCACCCACTCCCGTGACCGCCACGGCACATGCACACGCTCGCCCTCGCCCGCGG	251
Db	4623	CCCCCGCC	4564
Qy	252	CTTTTCTCGCTTCTACCGACCATGTGACCCCGCAGCAGCCTGCGCCAGCTGCCT	311
Db	4563	CCCCCGCC	4504
Qy	312	CCCGGGCAGTACTGGGGAC	330
Db	4503	CCCCCGCACTTCAAAGGAC	4485

RESULT 5

```

US-10-032-393-8/c
; Sequence 8, Application US/10032393
; Publication No. US20030027286A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Wall, Daniel
; APPLICANT: Gross, Molly
; TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
; FILE REFERENCE: ELITRA.010A
; CURRENT APPLICATION NUMBER: US/10/032.393
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/259,434
; PRIOR FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 09/948,993
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 12739
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Vector pPEF1
US-10-032-393-8

```

Matches 111; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 72 CCCCCCGTNNNGG 131
Db 4749 CC 4690

QY 132 GCCCCCCCCCGCAGGAGCTGTAGCCAGCCTCAGCCCGCAGCTCCAAAGAGAGCCCG 191
Db 4689 CC 4630

QY 192 GCAACCACTCCCGTGACCGCCAGCGCCACATGGACAGCCTGCCTCGCCTCGCCCGCG 251
Db 4629 CC 4570

QY 252 CTTTCTCTGCTTTCTACGAGCAGTGTGACCCGACACAGCCTGCCTCGCCTCGCCCT 311
Db 4569 CC 4510

QY 312 CCGGGCAGTACTGGGAC 330
Db 4509 CCCCCCACTTCAAGGAC 4491

RESULT 6
US-10-095-407-16
; Sequence 16, Application US/10095407
; Patent No. US20020164330A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/10/095,407
; PRIOR FILING DATE: 2002-03-11
; PRIOR APPLICATION NUMBER: US 60/091,650
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/054,646
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16:
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-10-095-407-16

Query Match 4.9%; Score 48.8; DB 9; Length 152331;
Best Local Similarity 52.5%; Pred. No. 0.15;
Matches 107; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 133 CCCCCCGCAGGAGCTGTAGCCCGCAGCCTCAGCCCGCAGCTCCAAAGAGAGCCCGG 192
Db 22019 CCCCCCGCCC 11

QY 193 CCACCACTCCCGTGACCCGCCAGCGCCACATGGACAGCCTGCCTCGCCTCGCCCGCG 252
Db 22079 CCCCCACCGAGCGCCCCCGCCCCCCCCCCCCCCCCCCCCCGCAGCGCCCCCGCCG 22138

QY 253 TTTTCTCTGCTTTCTACGAGCAGTGTGACCCGACAGCCTGCCTCGCCCGCAGCTC 312
Db 22139 CCGCCCGCAGCAGCCCCCGCCCCCGCGCGCCGCCGCCGCCGCCGCCGCCGCCGCC 22198

QY 313 CCGGGCAGTACTGGGACCTTCCC 336
Db 22199 CCCCCCGCGGGGCGCCCAACCC 22222

RESULT 7
US-10-078-090-83

; Sequence 83, Application US/10078090
; Publication No. US20030044815A1
; GENERAL INFORMATION:
; APPLICANT: Saiceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karra, Kalpana
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Gene
; FILE REFERENCE: DEX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; PRIOR FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 83
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-078-090-83

Query Match 4.8%; Score 48.2; DB 9; Length 954;
Best Local Similarity 55.8%; Pred. No. 0.015;
Matches 92; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 701 TATTAAATCTCGCTGTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 760
Db 70 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 129

QY 761 AATTTTCATTTTAAGCACATTTATCTAGAGAAATTTGCTGCTGATTCGGGGGAGCTG 820
Db 130 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 189

QY 821 GATCCAGAGCTGGAGGGGTGGTCCGGGGGAGGAGTGGCTCGG 865
Db 190 GCGCCCGCGGG 234

RESULT 8
US-10-012-542-44/c
; Sequence 44, Application US/10012542
; Publication No. US20030044851A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; CURRENT APPLICATION NUMBER: US/10/012,542
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-542-44


```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5517105"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT      189 a      354 c      288 g      236 t
ORIGIN
Query Match      71.5%; Score 715.8; DB 13; Length 1067;
Best Local Similarity 99.7%; Pred. No. 8e-112;
Matches 717; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 283 CCGCAGCAGCCCTGCCCCACCTCCCTCCGGCAGTACTGGGACCTTCCCTGGGG 342
Db 21 CCCTCACCAGCCCTGCCCCACCTCCCTCCGGCAGTACTGGGACCTTCCCTGGGG 80
QY 343 ACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402
Db 81 ACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 140
QY 403 GCTCCACAGCCTGGCTGAGCTCAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAG 462
Db 141 GCTCCACAGCCTGGCTGAGCTCAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAG 200
QY 463 CTTGGGTCTCAGGATGGTCTGGGGGCTGCTGTTCATCAAGACACCCCTGCGCCAG 522
Db 201 CTTGGGTCTCAGGATGGTCTGGGGGCTGCTGTTCATCAAGACACCCCTGCGCCAG 260
QY 523 TCACCACTCTTATCACCAGCAACGCCAGGACTTGGCTCCGCCATCCTCAGAACTCAC 582
Db 261 TCACCACTCTTATCACCAGCAACGCCAGGACTTGGCTCCGCCATCCTCAGAACTCAC 320
QY 583 AAGCATTGTCCTCCAGCTGGGGAACCTCAACCTCCGCCCTGCGTGGTGGTACAGAG 642
Db 321 AAGCATTGTCCTCCAGCTGGGGAACCTCAACCTCCGCCCTGCGTGGTGGTACAGAG 380
QY 643 GGGGTGGACAGGGGGGGGGTTCCTCTGTATACATCCCTGCGCATACACCCAGGTA 702
Db 381 GGGGTGGACAGGGGGGGGGTTCCTCTGTATACATCCCTGCGCATACACCCAGGTA 440
QY 703 TTAATTCCTGCTGTTTGTGTTTATTTATTTTATTTTATTTTATTTTATTTTATTAAGAA 762
Db 441 TTAATTCCTGCTGTTTGTGTTTATTTATTTTATTTTATTTTATTTTATTTTATTAAGAA 500
QY 763 TTTTCACTTTAAGACATTTATCTAAGGAATTTGTGCTGTATTTGGGGGAGCTGGA 822
Db 501 TTTTCACTTTAAGACATTTATCTAAGGAATTTGTGCTGTATTTGGGGGAGCTGGA 560
QY 823 TCAGAGCTGGAGGGGTGGTCCGGGGAGGAGTGGCTCGGAAGGGGCCCCCAGCTC 882
Db 561 TCAGAGCTGGAGGGGTGGTCCGGGGAGGAGTGGCTCGGAAGGGGCCCCCAGCTC 620
QY 883 CTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 942
Db 621 CTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 680
QY 943 TTAATACTGTGAAGTACTAATTTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1001
Db 681 TTAATACTGTGAAGTACTAATTTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 739
RESULT 2
BM927280      1091 bp      mRNA      linear      EST 12-MAR-2002
LOCUS
```

```
DEFINITION      AGENCOURT_668389 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5768363
5', mRNA sequence.
ACCESSION      BM927280
VERSION        BM927280.1 GI:19377659
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Primates; Catarrhini; Hominiidae; Homo.
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1091)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing By: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL12827 row: n column: 08
High quality sequence stop: 621.
```

FEATURES

```
Location/Qualifiers
1..1091
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5768363"
/clone_lib="NIH_MGC_121"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library."
BASE COUNT      278 a      339 c      286 g      188 t
ORIGIN
```

```
Query Match      69.2%; Score 692.6; DB 14; Length 1091;
Best Local Similarity 97.8%; Pred. No. 6.7e-108;
Matches 713; Conservative 0; Mismatches 14; Indels 2; Gaps 1;
QY 131 GGGCCCCCGCCAGGACGCTGTAGCCCGCAGCTCAGCCCCAGCTCCACAGAGCAGCC 190
Db 3 GGGCCCCCGCCAGGACGCTGTAGCCCGCAGCTCAGCCCCAGCTCCACAGAGCAGCC 62
QY 191 GGGCACCACCTCCCGTGACCGCCGACCATGACACAGCCCTGCGCCCTCCGCCCG 250
Db 63 GGGCACCACCTCCCGTGACCGCCGACCATGACACAGCCCTGCGCCCTCCGCCCG 122
QY 251 GCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 310
Db 123 GCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182
QY 311 TCCCGGGAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGT 370
Db 183 TCCCGGGAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGT 242
QY 371 GGACAGAGGCTGGGCGCTCAGTGGACTGCTGCCACAGCTGGGCTGACGTACAG 430
Db 243 GGACAGAGGCTGGGCGCTCAGTGGACTGCTGCCACAGCTGGGCTGACGTACAG 302
QY 431 GCGGAGGCCAGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 490
Db 303 GCGGAGGCCAGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 362
QY 491 CTGCTGTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAGC 550
```

```
Db 363 CTTGCTGTTTATCAAGAACCCCTCTGCCCCAGCTCACACACATCTTATCATCCAGCAACAGC 422
QY 551 CAGGACTTGGCTCCCTCCCTCCTCAGAACTCACAAGCATTGCTCCCTCAGCTGGGAACTT 610
Db 423 CAGGACTTGGCTCCCTCCCTCCTCAGAACTCACAAGCATTGCTCCCTCAGCTGGGAACTT 482
QY 611 CAACTTCCCTCCCTGCTCGCTGGTGGTGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCC 670
Db 483 CAACTTCCCTCCCTGCTCGCTGGTGGTGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCC 542
QY 671 CTGTACATACCTTGCATACCAACCCAGGTATTATTTCTCGCTGGTGGTGGTGGTGGTGGT 730
Db 543 CTGTACATACCTTGCATACCAACCCAGGTATTATTTCTCGCTGGTGGTGGTGGTGGTGGT 602
QY 731 TAAATTTTCTTTGTTGTTGTTTAAATAGAAATTTTCAATTTTAAAGCACATTTATATCTGAA 790
Db 603 TAAATTTTCTTTGTTGTTGTTTAAATAGAAATTTTCAATTTTAAAGCACATTTATATCTGAA 662
QY 791 GGAATTTCTGTGTTATTTGGGGG--AGCTGGATCCAGAGCTGGAGGGGGTGGTTCCTGG 848
Db 563 GGAATTTCTGTGTTATTTGGGGGGAATGGAATCCAAAGCTGGAGGGGGGGGACCTCC 722
QY 849 GGGAGGAG 857
Db 723 GAGCTAGAG 731

RESULT 3
BM554172
LOCUS
DEFINITION
AGENCOURT_6581002 NIH_MGC_41 Homo sapiens cdna clone IMAGE:5469285
5', mRNA sequence.
ACCESSION
BM554172
VERSION
BM554172.1 GI:18793541
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTP/DTP
cdna Library Preparation: Rubin Laboratory
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1974 row: d column: 22
High quality sequence stop: 593.
FEATURES
Location/Qualifiers
1..1160
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5469285"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dr priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT
173 a 324 c 280 g 219 t 164 others
ORIGIN
```

```
Query Match 67.6%; Score 676.4; DB 13; Length 1160;
Best Local Similarity 97.8%; Pred. No. 3.7e-105;
Matches 707; Conservative 0; Mismatches 12; Indels 4; Gaps 2;

QY 283 CCGCACCAGCCCTGCCCCACCTGCCCCAGCTGCTCCCGGCGAGTACTGGGAGCTTCCCTGGGG 342
Db 13 CTCGCACAGCCCTGCCCCACCTGCCCCAGCTGCTCCCGGCGAGTACTGGGAGCTTCCCTGGGG 72
QY 343 ACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402
Db 73 ACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 132
QY 403 GCTCCACAGCCCTGCCCGGCTGACGTCAGAGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGG 462
Db 133 GCTCCACAGCCCTGCCCGGCTGACGTCAGAGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAG 192
QY 463 CTTGGGCTCTCAGATGGGCTCCTGGGGGCTCGGTTCATCAAGACACCCCTCTGCCACAGC 522
Db 193 CTTGGGCTCTCAGATGGGCTCCTGGGGGCTCGGTTCATCAAGACACCCCTCTGCCACAGC 252
QY 523 TCACCATCTTTCATCACCAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 582
Db 253 TCACCATCTTTCATCACCAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 312
QY 583 AAGCCATTGCTCCCGGCTGGGGAACCTCAACCTCCCGGCTCGGTTCGTTGGTGACAGAG 642
Db 313 AAGCCATTGCTCCCGGCTGGGGAACCTCAACCTCCCGGCTCGGTTCGTTGGTGACAGAG 372
QY 643 GGGGTGGGACAGG-GGGGGGGGTTCCCGCTGTACATACCTGCCATACCAACCCAGGT 701
Db 373 GGGGTGGGACAGG-GGGGGGGGTTCCCGCTGTACATACCTGCCATACCAACCCAGGT 432
QY 702 ATTAATCTCGCTGCTGTTTGTGTTTATTTATTTTATTTTGTGTTTGTGTTTGTGTTTGTG 761
Db 433 ATTAATCTCGCTGCTGTTTGTGTTTATTTATTTTATTTTGTGTTTGTGTTTGTGTTTGTG 492
QY 762 ATTTTCATTTAAGCACATTTATCTAGAGGAATTTGCTGTGTATTTGGGGGAGCTGG 821
Db 493 ATTTTCATTTAAGCACATTTATCTAGAGGAATTTGCTGTGTATTTGGGGGAGCTGG 552
QY 822 ATCCAGAGCTGGAGGGGTGGTCCGGG---GAGGGAGTGGCTCGGAAGGGGGGGGGGGGG 878
Db 553 ATCCAGAGCTGGAGGGGTGGTCCGGG---GAGGGAGTGGCTCGGAAGGGGGGGGGGGGG 612
QY 879 TCTCCTTTCATGTCCTGTCGCCCGGCTGCTCCCTCCTCAGCCTTTTCTCCTCAGTTT 938
Db 613 TCTCCTTTCATGTCCTGTCGCCCGGCTGCTCCCTCCTCAGCCTTTTCTCCTCAGTTT 672
QY 939 CTCTTTAAACTGTGAAGTACTAACTTTCCAAAGGCTGCTCCCTCCTCCTCCCTCCCTCCCACTGGA 998
Db 673 CTCTTTAAACTGTGAAGTACTAACTTTCCAAAGGCTGCTCCCTCCTCCTCCCTCCCACTGGA 732
QY 999 GAA 1001
Db 733 GAA 735

RESULT 4
BM722189
LOCUS
DEFINITION
BM722189 UI-E-E00-ahy-m-11-0-UI.r1 UI-E-E00 Homo sapiens cdna clone
654 bp mRNA linear EST 01-MAR-2002
ACCESSION
BM722189 UI-E-E00-ahy-m-11-0-UI 5', mRNA sequence.
VERSION
BM722189.1 GI:19042595
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 654)
AUTHORS
Ronald M.F., Lennon, G. and Soares, M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
```



```

QY 131 GCGCCCCCGCCGAGCGAGCTGTAGCCCCAGCCCTCAGCCCCAGCTCCCAACAGAGCAGCCC 190
Db 655 GGGCCCCCCCCAGGCGAGCTGTAGCCCCAGCNCNTCAGCCCCAGCTCCCAACAGAGCAGCCC 596
QY 191 GCGCACCCACTCCCGTGTACCGCCCGCCAGCCACATGAGACAGACGCCCTCGCCCTCGGCCCGG 250
Db 595 GCGCACCCACTCNCCTGTACCGCCCGCCAGCCACATGAGACAGACGCCCTCGCCCTCGGCCCGG 536
QY 251 GCTTTCTCTGCTTCTTACCGACCATGTGACCCCGCCAGCCAGCCCTCGCCCGCCAGCCTGCC 310
Db 535 GCTTTCTCTGCTTCTTACCGACCATGTGACCCCGCCAGCCAGCCCTCGCCCGCCAGCCTGCC 476
QY 311 TCCGGGCGAGTACCTGGGAGCCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGG 370
Db 475 TCCGGGCGAGTACCTGGGAGCCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGG 416
QY 371 GGACAGAGGCGCTGGGCGCTCAGTGACCTGCTGCTGCCACAGCCCTGGGCTGACGTCAGAG 430
Db 415 GGACAGAGGCGCTGGGCGCTCAGTGACCTGCTGCTGCCACAGCCCTGGGCTGACGTCAGAG 356
QY 431 GCCAGGCGCAGAACTAGTGTAGGCGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 490
Db 355 GCCAGGCGCAGAACTAGTGTAGGCGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 296
QY 491 CTGCTGTTCATCAGACACCCCTCTGCCAGCTCAACCACTTTCATCACCAGCAAAAGC 550
Db 295 CTGCTGTTCATCAGACACCCCTCTGCCAGCTCAACCACTTTCATCACCAGCAAAAGC 236
QY 551 CAGGACTGTGCTCCCGCCTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTC 610
Db 235 CAGGACTGTGCTCCCGCCTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACT 176
QY 611 CAACCTCCCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 670
Db 175 CAACCTCCCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 116
QY 671 CTGTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 730
Db 115 CTGTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 56
QY 731 TAATTTTTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 785
Db 55 TAATTTTTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1

RESULT 6
BQ044936/c
LOCUS
DEFINITION
UI-H-EU0-azo-e-05-0-UI.s1 NCI-CGAP_Carl Homo sapiens cDNA clone
IMAGE:5851372 3', mRNA sequence.
BQ044936
ACCESSION
VERSION
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 654)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgap@rmail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
The following repetitive elements were found in this cDNA
sequence: 1-73, >AT rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD

```

FEATURES
Source
Location/Qualifiers
1..654

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5851372"
/tissue_type="Osteoarthritic Cartilage"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Knee; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site: 1: EcoR I; Site: 2: Not I; NCI-CGAP_Carl is a cDNA library containing the following tissue(s): Osteoarthritic Cartilage. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is TGATCAGCTT.

BASE COUNT 135 a 161 c 227 g 131 t
ORIGIN
TAG_LTB=UI-H-EU0
TAG_TISSUE=Osteoarthritic cartilage
TAG_SEQ=TGATCAGCTT

Query Match 63.9%; Score 639.6; DB 14; Length 654;
Best Local Similarity 98.6%; Pred. No. 7; le-99;
Matches 645; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 141 CCAGCAGCTGTAGCCCCAGCTCAGCCCCAGCTCAACAGAGCAGCGGCCACCCAC 200
Db 654 CCAGCAGCTGTAGCCCCAGCTCAGCCCCAGCTCAACAGAGCAGCGGCCACCCAC 595
QY 201 TCCCGTGTACCGCAGCATGTGACCCGCGCAGCCCTGCCCCAGCTGCCCTCCCGGGCAG 320
Db 594 TCCCGTGTACCGCAGCATGTGACCCGCGCAGCCCTGCCCCAGCTGCCCTCCCGGGCAG 475
QY 261 GCCTTTTACCGCAGCATGTGACCCGCGCAGCCCTGCCCCAGCTGCCCTCCCGGGCAG 380
Db 534 GCCTTTTACCGCAGCATGTGACCCGCGCAGCCCTGCCCCAGCTGCCCTCCCGGGCAG 415
QY 321 TACTGGGAGCCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 440
Db 474 TACTGGGAGCCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 355
QY 381 CTGGGCGCTCAGTGGAGCTGCTGCTGCCACAGCCTGGGCTGACGTCAGAGGCCGAGGCCA 500
Db 414 CTGGGCGCTCAGTGGAGCTGCTGCTGCCACAGCCTGGGCTGACGTCAGAGGCCGAGGCCA 295
QY 441 GGAAGTGTAGTGGGCCCTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 560
Db 354 GGAAGTGTAGTGGGCCCTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 235
QY 501 TCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAAAGCGGAGCTGG 620
Db 294 TCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAAAGCGGAGCTGG 175
QY 561 CTCGCCATCTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACT 680
Db 234 CTCGCCATCTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACT 115
QY 621 CTGCTCCTCGTGTGGTGACAGAGGGGTGGACAGAGGGGGGGGGTTCCTCCCTGTACATAC 740
Db 174 CTGCTCCTCGTGTGGTGACAGAGGGGTGGACAGAGGGGGGGGGTTCCTCCCTGTACATAC 55
QY 681 CTGCTCCTCAGCAACCCAGGATTAATTCCTCGCTGGTTTGTGTTTATTTATTTT 55
Db 114 CTGCTCCTCAGCAACCCAGGATTAATTCCTCGCTGGTTTGTGTTTATTTATTTT

```

QY 741 GTTTGATTTTAAAGAAATTTTCAATTTAAGACACATTTATATACTGAAGGAA 794
|||||
Db 54 GTTTGATTTTAAAGAAATTTTCAATTTAAGACACAAAAAATAAAAAA 1
|||||

RESULT 7
BQ441751 954 bp mRNA linear EST 24-MAY-2002
AGENCOURT_7761844 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6018014
5'; mRNA sequence.
ACCESSION BQ441751
VERSION BQ441751.1 GI:21180827
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 954)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13218 row: d column: 15
High quality sequence stop: 563.
Location/Qualifiers
1..954
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6018014"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 149 a 298 C 271 g 235 t 1 others
ORIGIN
Query Match 62.8%; Score 628.6; DB 14; Length 954;
Best Local Similarity 98.4%; Pred. No. 4.8e-97;
Matches 634; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 309 CCTCCGGGCGAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGACTCC 368
|||||
Db 1 CCTCCGGGCGAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGACTCC 60
|||||

QY 369 TTGACAGAGGCTGGGCGCTCAGTGACCTGCCTGCCACAGCCCTGGGCTGACGTCAG 428
|||||
Db 61 TTGACAGAGGCTGGGCGCTCAGTGACCTGCCTGCCACAGCCCTGGGCTGACGTCAG 120
|||||

QY 429 AGGCCAGGCCAGGAACATGAGTAGGAGCCCTGGTCTGGTCTCAGGATGGGCTCGGGG 488
|||||
Db 121 AGGCCAGGCCAGGAACATGAGTAGGAGCCCTGGTCTGGTCTCAGGATGGGCTCGGGG 180
|||||

QY 489 GCCTCTGTTCATCAGACACACCCCTCTGCCAGCTACACATCTTCATCACCAGCAAC 548
|||||
Db 181 GCCTCTGTTCATCAGACACACCCCTCTGCCAGCTACACATCTTCATCACCAGCAAC 240
|||||

QY 549 GCAGGACTTGGTCTCCCCCTCTCAGAACTCAGAACCCATTGCTCCCCAGCTGGGGAAC 608
|||||
Db 241 GCAGGACTTGGTCTCCCCCTCTCAGAACTCAGAACCCATTGCTCCCCAGCTGGGGAAC 300
|||||

QY 609 CTCACACTCTCCCTCGCTGGTGTGGTGACAGAGGGGGTGGGACAGGGGGGGGGTTC 668
|||||

```

```

Db 301 CTCAACTCCCTCCCTCGCTGGTGTGGACAGAGGGGTGGACAGGGGGGGGTCC 360
|||||
QY 669 CCTGTACATACCTCCATACCAACCCAGGTATTAATCTCGCTGGTGTGGTGTAT 728
|||||
Db 361 CCTGTACATACCTCCATACCAACCCAGGTATTAATCTCGCTGGTGTGGTGTAT 420
|||||
QY 729 TTAAATTTTGTGTTTGTATTTTAAAGAAATTTTCAATTTAAGACACATTTACTG 788
|||||
Db 421 TTAAATTTTGTGTTTGTATTTTAAAGAAATTTTCAATTTAAGACACATTTACTG 480
|||||
QY 789 AAGGAATTTGTCTGTATTTGGGGAGCTGCATCAGAGCTGGAGGGGGTGGTCCGG 848
|||||
Db 481 AAGGAATTTGTCTGTATTTGGGGAGCTGCATCAGAGCTGGAGGGGGTGGTCCGG 540
|||||
QY 849 GGGAGGAGTGGCTCGGAAGGGGCCCCACTCTCTCTTCATGCTCCCTGTCCTCCCACTGTT 908
|||||
Db 541 GGGAGGAGTGGCTCGGAAGGGGCCCCACTCTCTCTTCATGCTCCCTGTCCTCCCACTGTT 600
|||||
QY 909 CTCCTCTCAGCTTTTCTCTCTCAGTCTTCTCTTAAACTGT 952
|||||
Db 601 CGGCTGCTCAGCTTTTCTCTCTCAGTCTTCTCTTATAACTGT 644
|||||

RESULT 8
AL522712 978 bp mRNA linear EST 13-FEB-2001
LOCUS AL522712 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB009YG04 3
DEFINITION prime, mRNA sequence.
ACCESSION AL522712
VERSION AL522712.1 GI:12786205
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 978)
L.I.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..978
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB009YG04"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6: 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fllang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 167 a 293 c 331 g 177 t 10 others
ORIGIN
Query Match 60.28; Score 602.6; DB 9; Length 978;
Best Local Similarity 90.9%; Pred. No. 1.2e-92;
Matches 613; Conservative 5; Mismatches 55; Indels 1; Gaps 1;

QY 77 CGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136
|||||
Db 673 CCTGGACACTCTGACCGACAGCGGGGGTGGGGCGGACGCGGGGTGCGCTGGCCCC 614
|||||

```

```
QY 137 CCGCCAGGAGCTGTAGCCCGAGCCTCAGCCCGAGCTCCACAGAGAGCCCGGCGAC 196
|||||
Db 613 CCGCCAGGAGCTGTAGCCCGAGCCTCAGCCCGAGCTCCACAGAGAGCCCGGCGAC 554
QY 197 CCACCTCCCGGTACCGCCCGCCAGCCACATGACACAGCCCTCGCCCTCCCGCCCGGCTTTT 256
|||||
Db 553 CCACCTCCCGGTACCGCCCGCCAGCCACATGACACAGCCCTCGCCCTCCCGCCCGGCTTTT 494
QY 257 CTCCTGCTTTCACGACCATGTGACCCCGCAGCCCTGCTCCCGCCAGCTGCTCCCGG 316
|||||
Db 493 CTCCTGCTTTCACGACCATGTGACCCCGCAGCCCTGCTCCCGCCAGCTGCTCCCGG 435
QY 317 GCAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376
|||||
Db 434 GCAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 375
QY 377 AGGCTGGGCGCTCAGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 436
|||||
Db 374 AGGCTGGGCGCTCAGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 315
QY 437 GCCAGGAATCAGTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 496
|||||
Db 314 GCCAGGAATCAGTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 255
QY 497 TTCATCAGACACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 556
|||||
Db 254 TTCATCAGACACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 195
QY 557 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 616
|||||
Db 194 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 135
QY 617 CCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 676
|||||
Db 134 CCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 75
QY 677 ATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 736
|||||
Db 74 ATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15
QY 737 TTTTGTGTTTGAATTT 750
|||||
Db 14 TTTTGTGTTTGAATTT 1

RESULT 9
AL533351
LOCUS
DEFINITION
AL533351 Lri_Fl015_Brn1 Homo sapiens cDNA clone CS0DN004Y122 5
prime, mRNA sequence.
ACCESSION
AL533351
VERSION
AL533351.1 GI:12796844
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 967)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
CONTACT
Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DN004Y122"
/clone_lib="Lri_Fl015_Brn1"
/sex="male"
```

```
/tissue_type="Adult brain"
/note="Vector: pcMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pcMVSPORT 6
vector. Library was constructed by Life Technologies.
Contact : Feng Liang Life Technologies, a division of
Invitrogen 9800 Medical Center Drive Rockville, Maryland
20850, USA Fax : (1) 301 610 8371 Email :
filiang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 173 a 325 c 286 g 177 t 6 others
ORIGIN
```

```
Query Match 59.8%; Score 599; DB 9; Length 967;
Best Local Similarity 90.9%; Pred. No. 4.9e-92;
Matches 621; Conservative 5; Mismatches 55; Indels 2; Gaps 2;
```

```
QY 77 CGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136
|||||
Db 285 CTTGGACACTCTGAGCGGACAGCGCGGGGGTGGGGGGGCGGAGCGGGGTGGCTGGCC 344
QY 137 CCGCGCCAGGAGCTGTAGCCCGCCAGCCTCAGCCCGCCAGCTCCACAGAGAGCGCCGCGCAC 196
|||||
Db 345 CCGCGCCAGGAGCTGTAGCCCGCCAGCCTCAGCCCGCCAGCTCCACAGAGAGCGCCGCGCAC 404
QY 197 CCACCTCCCGGTGACCGCCCGCCAGCCACATGACACAGCCCTCGCCCTCGCCCGGCTTTT 256
|||||
Db 405 CCACCTCCCGGTGACCGCCCGCCAGCCACATGACACAGCCCTCGCCCTCGCCCGGCTTTT 464
QY 257 CTCCTGCTTTCACGACCATGTGACCCCGCCAGCCTGCTGCTGCTGCTGCTGCTGCTGCT 316
|||||
Db 465 CTCCTGCTTTCACGACCATGTGACCCCGCCAGCCTGCTGCTGCTGCTGCTGCTGCTGCT 524
QY 317 GCAGTACTGGGAGCCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376
|||||
Db 525 GCAGTACTGGGAGCCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 584
QY 377 AGGCTGGGCGCTCAGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 436
|||||
Db 585 AGGCTGGGCGCTCAGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 644
QY 437 GCCAGGAATCAGTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 495
|||||
Db 645 GCCAGGAATCAGTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 704
QY 496 GTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGCCAGGA 555
|||||
Db 705 GTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGCCAGGA 764
QY 556 CTTGGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACTCAAC 615
|||||
Db 765 CTTGGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACTCAAC 824
QY 616 TCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 674
|||||
Db 825 TCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
QY 675 ACATACCTGCCATCACCACCCAGGATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 734
|||||
Db 885 ACATACCTGCCATCACCACCCAGGATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
QY 735 TTTTGTGTTTGAATTTTAAAT 757
|||||
Db 945 TTTTGTGTTTGAATTTTAAAT 967
```

```
RESULT 10
AL531149 Lri_NFL001_NBC4 Homo sapiens cDNA clone CS0DN001Y13 3
LOCUS prime, mRNA sequence.
DEFINITION
AL531149 Lri_NFL001_NBC4 Homo sapiens cDNA clone CS0DN001Y13 3
ACCESSION
AL531149
```

[illegible]

QY	288	CTCCATTGAGACCCAGACGACAGTTCCTGAAGAGATAGTGCCAGGCCCTCCCTCGCCACC	347
DB	279	CACCATTGAGACCCAGACGACGAGTCTGAGAGATFAGTGCCAGGCCCTCCCTCGCCACC	338
QY	348	CCCTCTACCCCGCATCTACAAGCCTTGCTTTGTCTGTGTCAGGACAAGTTCCTCAGGCTACCA	407
DB	339	CCCTCTACCCCGCATCTACAAGCCTTGCTTTGTCTGTGTCAGGACAAGTTCCTCAGGCTACCA	398
QY	408	CTATGGGTTCAGCGCCTGTGAGGGCTCGAAGGCTCTTCCGCCGAGCATCCAGAAGAA	467
DB	399	CTATGGGTTCAGCGCCTGTGAGGGCTCGAAGGCTCTTCCGCCGAGCATCCAGAAGAA	458
QY	468	CATGTTGTACACGTGTCAACCGGACAAGAACTGCATCATCAACAAGGTCAACCGGACCC	527
DB	459	CATGTTGTACACGTGTCAACCGGACAAGAACTGCATCATCAACAAGGTCAACCGGACCG	518
QY	528	CTGCCAGTACTGCCGACTCGAGAGTGCCTTTGAAGTGGGCATGTCACAAGAGTCTGTGTAG	587
DB	519	CTGCCAGTACTGCCGACTCGAGAGTGCCTTTGAAGTGGGCATGTCACAAGAGTCTGTGTAG	578
QY	588	AAACGACCGAAACAAGAAAGAGAGAGGTGCCCAAGCCGGAGTGCCTGTGAGAGTACAC	647
DB	579	AAACGACCGAAACAAGAAAGAGAGAGGTGCCCAAGCCGGAGTGCCTGTGAGAGTACAC	638
QY	648	GCTGACCCGGAGTGGGGAGTCTCAATTGAGAAGTGGCCAAGCGCACCAAGAAACCTT	707
DB	639	GCTGACCCGGAGTGGGGAGTCTCAATTGAGAAGTGGCCAAGCGCACCAAGAAACCTT	698
QY	708	CCCTGCCCTTGGCAGTGGGCAAAATACACTAGAACAAACAGCTCAGAACAAACGTGTCTC	767
DB	699	CCCTGCCCTTGGCAGTGGGCAAAATACACTAGAACAAACAGCTCAGAACAAACGTGTCTC	758
QY	768	TCTGGACATTGACCTCTGGGACAAGTTCAGTGAACCTCCACCAAGTGCATTTAAGAC	827
DB	759	TCTGGACATTGACCTCTGGGACAAGTTCAGTGAACCTCCACCAAGTGCATTTAAGAC	818
QY	828	TGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAACACCTCACCATCGCCGACAGATCAC	887
DB	819	TGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAACACCTCACCATCGCCGACAGATCAC	878
QY	888	CTCTCTCAAGGCTGCCTGGACATCTGTATCTGTGGGATCTGCACGCGGTACACGCC	947
DB	879	CTCTCTCAAGGCTGCCTGGACATCTGTATCTGTGGGATCTGCACGCGGTACACGCC	938
QY	948	CGACGAGACACCATGACCTTCTCGACGGGCTGACCTGAACCGGACCCAGATGCACAA	1007
DB	939	CGACGAGACACCATGACCTTCTCGACGGGCTGACCTGAACCGGACCCAGATGCACAA	998
QY	1008	CGTGGCTTCGGCCCCCTCACCGACCTGGTCTTTTGCTTCGCCAACACAGTGTGCCCT	1067
DB	999	CGTGGCTTCGGCCCCCTCACCGACCTGGTCTTTTGCTTCGCCAACACAGTGTGCCCT	1058
QY	1068	GGAGATGGATGATGCGGAGACGGGGCTGCTACGCGCCATCTGCCTCATPTCGCGAGACCG	1127
DB	1059	GGAGATGGATGATGCGGAGACGGGGCTGCTACGCGCCATCTGCCTCATPTCGCGAGACCG	1118
QY	1128	CCAGGACCTTGGAGACGCGGACCGGAGTGTCTGAGAGCGGCTGCTGAGAGCGCT	1187
DB	1119	CCAGGACCTTGGAGACGCGGAGTGTCTGAGAGCGGCTGCTGAGAGCGCT	1178
QY	1188	AAAGTCTTACGTGCGGAAGCGGAGGCCAGCGCCCCCACATGTTCCCAAGATGCTTAAT	1247
DB	1179	AAAGTCTTACGTGCGGAAGCGGAGGCCAGCGCCCCCACATGTTCCCAAGATGCTTAAT	1238
QY	1248	GAAGATTACTGTGCGAAGCATCAGCGCAAGGGGCTGAGCGGTGATCAGCTGAA	1307
DB	1239	GAAGATTACTGTGCGAAGCATCAGCGCAAGGGGCTGAGCGGTGATCAGCTGAA	1298
QY	1308	GATGGAGATCCCGGGCTCCATCCCGCCTCTCATCCAGAAATGTTGAGAACTCAGAGG	1367
DB	1299	GATGGAGATCCCGGGCTCCATCCCGCCTCTCATCCAGAAATGTTGAGAACTCAGAGG	1358

RESULT 2

US-09-797-727-1

; Sequence 1, Application US/09797727

Patent No. US20020077457A1

: RECEIVED NO: 032002007
: GENERAL INFORMATION:

GENERAL INFORMATION:

APPLICANT: The Salk Institute

APPLICANT: TAKAKU, Fumimaro

; TITLE OF INVENTION: GAMMA RETINOIC ACID RECEPTOR

; FILE REFERENCE: SALK1150-3

: CURRENT APPLICATION NUMBER: US/0

CURRENT REFLECTION NUMBER: 2001-08-31

; CURRENT FILING DATE: 2001-08-31
 : PRIOR APPLICATION NUMBER: US 00/400 200

; PRIOR APPLICATION NUMBER: US 08/486,325
 ; PRIOR PCT NUMBER: PCT/US 95/00000

;
PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: US 08/100,039

; PRIOR FILING DATE: 1993-07-30

; PRIOR APPLICATION NUMBER: PCT/US90/0356

: PRIOR FILING DATE: 1990-06-22

; PRIOR FILING DATE: 1990
 . PRIOR APPLICATION NUMBER:

;; PRIOR APPLICATION NUMBER: US 0
: PRIOR FILING DATE: 2008-06-03

;
PRIOR FILLING

; NUMBER OF SEQ

; SOFTWARE:

[illegible]

ORGANISM: Artificial Sequence

FEATURE: Description of Artificial Sequence:

OTHER INFORMATION: PONY-FLAG-RARBeta2 vector genome plasmid

US-10-239-804-10

Query Match 33.6%; Score 701.2; DB 9; Length 9151;
Best Local Similarity 77.4%; Pred. No. 4.6e-169;
Matches 850; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

QY 286 CACTCCATTGAGACCCAGACGAGTCTCTGAAGATAGTCCCGAGCCCTCCCTCGCCA 345
DB 3018 CAATCATCGAGACACAGATACCAAGCTCTGAGAGCTCGTCCCGAGCCCATCTCCA 3077
QY 346 CCCCCCTACCCCGCATCTACAAAGCTTGTCTGTGAGGACAAAGTCCCTCAGCTAC 405
DB 3078 CTTCCTCTCCTCGGTGTACAAAGCTTGTCTGTGAGGACAAAGTCCCTCAGCTAC 3137
QY 406 CACTATGGGTGAGCGCTGTGAGGGTGTCAAGGGCTTCTCCGCGCAGCATCCAGAAG 465
DB 3138 CACTATGGGTGAGCGCTGTGAGGGTGTCAAGGGCTTCTCCGCGCAGCATCCAGAAG 3197
QY 466 AACATGGTGTACACCTGTACCGGACAAAGTCAATCAACAAAGTGAACCGGAAC 525
DB 3198 AACATGGTGTACACCTGTACCGGACAAAGTCAATCAACAAAGTGAACCGGAAC 3257
QY 526 CACTCCAGTACTGCGAGTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 585
DB 3258 CACTCCAGTACTGCGAGTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3317
QY 586 AGAAACGACCGAAGAAAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 645
DB 3318 AGAAACGACCGAAGAAAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3377
QY 646 ACGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 705
DB 3378 ACGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3437
QY 706 TTCCCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 765
DB 3438 TTCCCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3497
QY 766 TCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 825
DB 3498 TCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3557
QY 826 ACTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 885
DB 3558 ACTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3617
QY 886 ACCCTCTCAAGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 945
DB 3618 ACCCTCTCAAGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3677
QY 946 CCGGAGGAGGACCATGACCTTCTGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAG 1005
DB 3678 CCGGAGGAGGACCATGACCTTCTGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAG 3737
QY 1006 AACGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1065
DB 3738 AACGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3797
QY 1066 CTGAGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1125
DB 3798 CTGAGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3857
QY 1126 CGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1185
DB 3858 CGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3917
QY 1186 CTAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1245
DB 3918 CTAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3977

QY 1246 ATGAAGATTACTGACCTGCGAAGCATCAGCGCAAGGGGCTGAGCGGGTGCATCACGCTG 1305
DB 3978 ATGAAGATTACTGACCTGCGAAGCATCAGCGCAAGGGGCTGAGCGGGTGCATCACGCTG 4037
QY 1306 AGATGAGATCCCGGGTCCATGCGGCTCTCATCCAGGAATGTTGGAGAACTCAGAG 1365
DB 4038 AATATGAAATTCCTGGATCAATGCCACCTCTCATTCAGGAATGCTGGAGAAATCTGAA 4097
QY 1366 GGCCTGGACACTCTGAGC 1383
DB 4098 GGACATGAACCTTGACC 4115

RESULT 7

US-09-822-830A-614/C
; Sequence 614, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 614
; LENGTH: 595
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-614

Query Match 28.4%; Score 591.4; DB 10; Length 595;
Best Local Similarity 99.8%; Pred. No. 1.3e-141;
Matches 592; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1492 TCCCGTGTACCGCCACGACCATGTGACAGCCCTGCGCCCTCGCCCGCGCTTTCTCT 1551
DB 593 TCCCGTGTACCGCCACGACCATGTGACAGCCCTGCGCCCTCGCCCGCGCTTTCTCT 534
QY 1552 GCCTTTCTACCGACCATGTGACCCCGCAGCCCTGCCCCACCTGCCCTCCCGGGCAG 1611
DB 533 GCCTTTCTACCGACCATGTGACCCCGCAGCCCTGCCCCACCTGCCCTCCCGGGCAG 474
QY 1612 TACTGGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1671
DB 473 TACTGGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 414
QY 1672 CTGGGCGCTCAGTGGAGTGCCTGCCACAGCTGGGCTGAGGTGAGGTGAGGTGAGGTGAG 1731
DB 413 CTGGGCGCTCAGTGGAGTGCCTGCCACAGCTGGGCTGAGGTGAGGTGAGGTGAGGTGAG 354
QY 1732 GGAACGTGAGTGGGCGCTTGGTCTCAGGATGGTCTGAGGATGGTCTGAGGATGGTCTGAG 1791
DB 353 GGAACGTGAGTGGGCGCTTGGTCTCAGGATGGTCTGAGGATGGTCTGAGGATGGTCTGAG 294
QY 1792 TCAAGACACCCCTCTGCCAGCTCACCACATCTTATCACCAGCAAAACGACGAGTGTG 1851
DB 293 TCAAGACACCCCTCTGCCAGCTCACCACATCTTATCACCAGCAAAACGACGAGTGTG 234
QY 1852 CTCGCCCGCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGAGAACTCAACCTCCCC 1911
DB 233 CTCGCCCGCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGAGAACTCAACCTCCCC 174

QY 1912 CCTGCTCGGTTGGTGACAGAGGGGTGGACAGGGGGGGGGTCCCTGTACATAC 1971
Db 173 CCTGCTCGGTTGGTGACAGAGGGGTGGACAGGGGGGGGGTCCCTGTACATAC 114
QY 1972 CCTGCCATACCAACCCAGGATTAATCTCGCTGGTTGTTTATTTAATTTT 2031
Db 113 CCTGCCATACCAACCCAGGATTAATCTCGCTGGTTGTTTATTTAATTTT 54
QY 2032 GTTTTGATTTTTTAATAAGAAATTTCAATTTAAGCAAAAAA 2084
Db 53 GTTTTGATTTTTTAATAAGAAATTTCAATTTAAGCAAAAAA 1
RESULT 8
US-09-960-352-5714
; Sequence 5714, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5714
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 25-LIB188-001-Q1-E1-G1
US-09-960-352-5714

Query Match 18.0%; Score 375.6; DB 10; Length 411;
Best Local Similarity 95.3%; Pred. No. 1.9e-86;
Matches 387; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 536 ACTGCCGACTGCAGAAAGTCTTGAAGTGGGATGTCCAAAGGAGTCTGTGAGAAACGACC 595
Db 1 ACTGCCGACTGCAGAAAGTCTTGAAGTGGGATGTCCAAAGGAGTCTGTGAGAAATGACC 60
QY 596 GAAACAAGAGAAGAGAGGTGCCAAGCCGAGTGTCTGAGAGCTACACGCTGACGC 655
Db 61 GGAACAGGAAAAAGAGAGGTGCCAAGCCGAGTGTCTGAGAGCTACACGCTGACGC 120
QY 656 CGGAGGTGGGGAGCTCATTTGAGAGGTGGCAAGCCGACCAAGAAACCTTCCCTGCC 715
Db 121 CGGAGGTGGGGAGCTCATTTGAGAGGTGGCAAGCCGACCAAGAAACCTTCCCTGCC 180
QY 716 TCTGCCAGCTGGGCAAAATACACTACGAAACAGCTCAGAAACAGTGTCTCTGTGACA 775
Db 181 TCTGCCAGCTGGGCAAAATACACTACGAAACAGCTCAGAAACAGTGTCTCTGTGACA 240
QY 776 TTGACCTCTGGGACAAGTTCAGTGAATCTCCCAAGTGCATCATTAAGACTGTGGAGT 835
Db 241 TTGACCTCTGGGACAAGTTCAGTGAATCTCCCAAGTGCATCATTAAGACTGTGGAGT 300
QY 836 TCGCCAAGCAGCTCCCGGTTTACACCTTCCACCTCAGCAGCAGATCACCTCTCTCA 895
Db 301 TCGCCAAGCAGCTCCCGGTTTACACCTTCCACCTCAGCAGCAGATCACCTCTCTCA 360
QY 896 AGGCTGCTGCTGGAGATCTGTATCTCTGGATCTGCACCGGTA 941
Db 361 AGGCTGCTGCTGGAGATCTGTATCTCTGGATCTGCACCGGTA 406

RESULT 9
US-09-960-352-12806/c
; Sequence 12806, Application US/09960352
; Patent No. US20020137139A1

; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12806
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 55-LIB188-005-Q1-E1-F4
US-09-960-352-12806

Query Match 18.0%; Score 375.6; DB 10; Length 417;
Best Local Similarity 95.3%; Pred. No. 2e-86;
Matches 387; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 536 ACTGCCGACTGCAGAAAGTCTTGAAGTGGGATGTCCAAAGGAGTCTGTGAGAAACGACC 595
Db 417 ACTGCCGACTGCAGAAAGTCTTGAAGTGGGATGTCCAAAGGAGTCTGTGAGAAATGACC 358
QY 596 GAAACAAGAGAAGAGAGGTGCCAAGCCGAGTGTCTGAGAGCTACACGCTGACGC 655
Db 357 GGAACAAGAAAAAGAGAGGTGCCAAGCCGAGTGTCCGAGAGCTACACGCTGACGC 298
QY 656 CGGAGGTGGGGAGCTCATTTGAGAGGTGGCAAGCCGACCAAGAAACCTTCCCTGCC 715
Db 297 CGGAGGTGGGGAGCTCATTTGAGAGGTGGCAAGCCGAGTGTCCCTGCTC 238
QY 716 TCTGCCAGCTGGGCAAAATACACTACGAAACAGCTCAGAAACAGTGTCTCTGTGACA 775
Db 237 TCTGCCAGCTGGGCAAAATACACTACGAAACAGCTCAGAAACAGTGTCTCTGTGACA 178
QY 776 TTGACCTCTGGGACAAGTTCAGTGAATCTCCCAAGTGCATCATTAAGACTGTGGAGT 835
Db 177 TTGACCTCTGGGACAAGTTCAGTGAATCTCCCAAGTGCATCATTAAGACTGTGGAGT 118
QY 836 TCGCCAAGCAGCTCCCGGTTTACACCTTCCACCTCAGCAGCAGATCACCTCTCTCA 895
Db 117 TCGCCAAGCAGCTCCCGGTTTACACCTTCCACCTCAGCAGCAGATCACCTCTCTCA 58
QY 896 AGGCTGCTGCTGGAGATCTGTATCTCTGGATCTGCACCGGTA 941
Db 57 AGGCTGCTGCTGGATCTGTATCTCTGGATCTGCACCGGTA 12

RESULT 10
US-09-960-352-31
; Sequence 31, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 31
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 01-LIB188-009-Q1-E1-A1
US-09-960-352-31

US-09-960-352-2138
; Sequence 2138, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagapan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 2138
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (76)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 10-LIB188-020-Q1-E1-C5
US-09-960-352-2138

Query Match 13.5%; Score 282.4; DB 10; Length 416;
Best Local Similarity 94.3%; Pred. No. 1.4e-62;
Matches 314; Conservative 0; Mismatches 17; Indels 2; Gaps 2;

QY	613	GAGTGTCCCAAGCCCGAGTGCTCTGAGAGCTACACGCTGA-CGCCGGAGGTGGGGGAGCT	671
Db	12	GGGATACCAAGCAGCAGTGCTCCGAGAGCTACACGTTGACCCCGGAGGTGGGGGAGCT	71
QY	672	CAT-TGAGAGGTGCGCAAGCGCACGAGAACTTCCCTGCCCTCTGCCAGCTGGGCA	730
Db	72	CATCNGASAGGTGCGCAAGCGCATCAGAGAGCTTCCCTGCTCTGCCAAGCTGGGCA	131
QY	731	AATACACTAGCAACACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACA	790
Db	132	AATACACTAGCAACACAGCTCAGAACAGCGTGCTCTCTGGACATTGACCTCTGGGACA	191
QY	791	AGTTCAGTGAACCTCCACCAAGTGATCATTAAGACTGTGGAGTTCGCCAAGCACTGC	850
Db	192	AGTTCAGTGAACCTCCACCAAGTGATCATTAAGACTGTGGAGTTCGCCAAGCACTGC	251
QY	851	CCGGCTTCACACCCCTCACCATCGCCGACAGATCACCCCTCCTCAAGGCTGCCTGCTGG	910
Db	252	CCGGCTTCACACCCCTCACCATTGCGGACAGATCACCCCTCCTCAAGGCTGCCTGCTGG	311
QY	911	ACATCTGTATCTCGGGATCTGCACGCGGTACA	943
Db	312	ATATCTGTATCTCGGGATCTGCACGCGGTACA	344

RESULT 14
US-09-962-436-567
; Sequence 567, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 567
; LENGTH: 2288

[illegible]

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 09:48:44 ; Search time 2310 Seconds
(without alignments)
14625.015 Million cell updates/sec

Title: US-09-691-220-1

Perfect score: 2086

Sequence: 1 aacgacgagggcgagggg.....cacaaaaa 2086

Scoring table: IDENTITY_NUC

Searched: Gapop 10.0 , Gapext 1.0

16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	960.8	46.1	1051	9	AL578663	AL578663
C 2	923.6	44.3	967	9	AL533351	AL533351
C 3	923.4	44.3	978	9	AL522712	AL522712
C 4	903.6	43.3	992	9	AL556392	AL556392
C 5	882.2	42.3	931	9	AL531149	AL531149
C 6	842.8	40.4	1073	13	BM544324	BM544324

C 7	807.2	38.7	993	9	AL551103	AL551103
C 8	729.6	35.0	885	14	BQ706025	BQ706025
C 9	678.8	32.5	3141	11	BC030234	BC030234
C 10	659.8	31.6	702	13	BI827961	BI827961
C 11	654	31.4	654	14	BQ044936	BQ044936
C 12	648.4	31.1	1091	14	BM927280	BM927280
C 13	645.4	30.9	703	9	AI806984	AI806984
C 14	639	30.6	1164	12	BG028346	BG028346
C 15	635	30.4	646	14	BM790595	BM790595
C 16	633.8	30.4	700	13	BI907041	BI907041
C 17	623	29.9	692	9	AA812217	AA812217
C 18	620.2	29.7	812	13	BI914043	BI914043
C 19	604.2	29.0	861	10	BE547412	BE547412
C 20	599.4	28.7	655	12	BG818232	BG818232
C 21	597.8	28.7	626	9	AI963441	AI963441
C 22	585	28.0	1123	13	BM468687	BM468687
C 23	572.8	27.5	602	10	AW118835	AW118835
C 24	569.8	27.3	764	9	AA790328	AA790328
C 25	566.8	27.2	670	12	BG337888	BG337888
C 26	566.8	27.2	1079	14	BM909949	BM909949
C 27	562.8	27.0	585	13	BI770450	BI770450
C 28	560	26.8	560	14	BM817785	BM817785
C 29	559.2	26.8	596	9	AI962618	AI962618
C 30	559.2	26.8	982	12	BE794377	BE794377
C 31	550.4	26.4	866	12	BF182871	BF182871
C 32	546.2	26.2	552	9	AI380422	AI380422
C 33	540.4	25.9	1034	12	BE871903	BE871903
C 34	529	25.4	743	13	BI830768	BI830768
C 35	513	24.6	885	13	BI663939	BI663939
C 36	504.8	24.2	911	9	AL526754	AL526754
C 37	502.8	24.1	583	9	AI401587	AI401587
C 38	495	23.7	954	14	BQ951520	BQ951520
C 39	494.4	23.7	1067	13	BM423874	BM423874
C 40	492	23.6	921	14	BQ714956	BQ714956
C 41	488.2	23.4	564	10	AV609965	AV609965
C 42	484.6	23.2	782	13	BI856804	BI856804
C 43	483.4	23.2	1061	14	BQ278797	BQ278797
C 44	483.4	23.2	1160	13	BM554172	BM554172
C 45	483	23.2	483	9	AI640571	AI640571

ALIGNMENTS

RESULT 1
AL578663/C
LOCUS
DEFINITION
AL578663 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK004YG05 3
prime, mRNA sequence.
ACCESSION
AL578663
VERSION
AL578663.1 GI:12942954
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1051)
AUTHORS
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..1051
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK004YG05"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: PCWSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 189 a 311 c 354 g 189 t 8 others
ORIGIN

Query Match 46.1%; Score 960.8; DB 9; Length 1051;
Best Local Similarity 97.2%; Pred. No. 8.5e-127;
Matches 1023; Conservative 7; Mismatches 17; Indels 6; Gaps 5;

QY 992 GGACCCAGATGCACACGCTGCTCGGCCCTCACCACCTGCTCTTTGGCTTCGCCA 1051
DB 1051 GGACCCAGATGCACACGCTGCGC-TYGGCCCTTCACCGACTGCTTTGGCTTCGCCA 993
QY 1052 ACCAGCTGCTGCCCTGGAGATGATGATGCGAGACGGGGTGTCTCAGCGCATCTGCC 1111
DB 992 ACCAGCTGCTGCCCTGGAGATGATGATGCGAGACGGGGTGTCTCAGCGCATCTGCC 993
QY 1112 TCATCTGGGAGACCGCAGGACCTGGAGCA-GCCGGACCGGGTGGACATGTGCGAGG 1170
DB 932 TCATCTGGGAGACCGCAGGACCTGGAGCAGCGCGGGTGGACATGTGCGAGGAG 873
QY 1171 CGCGTCTGGAGGCGCT-AAAGGTCTAGTGGGGAAGCGGAGCGCCAGCGCCCGC -AC 1227
DB 872 CGCGTCTGGAGGCGCTAAAGGTCTAGTGGGGAAGCGGAGCGCCAGCGCCCGCACH 813
QY 1228 ATGTTCCCAAGATGCTAATGAAGATTACTGACCTGCGAAGCATCAGCGCCAGGGGGT 1287
DB 812 ATGTTCCCAAGATGCTAATGAAGATTACTGACCTGCGAAGCATCAGCGCCAGGGGGT 753
QY 1288 GAGCGGTGATCACCTGAAGATGAGATCCGGGCTCATGCCGCTCTCATCCAGGAA 1347
DB 752 GAGCGGTGATCACCTGAAGATGAGATCCGGGCTCATGCCGCTCTCATCCAGGAA 693
QY 1348 ATGTTGGAGAACTCAGAGGGCTGACACTCTGACGGACACCGGGGGTGGGGGGCG 1407
DB 692 ATGTTGGAGAACTCAGAGGGCTGACACTCTGACGGACACCGGGGGTGGGGGGCGG 633
QY 1408 GACGGGTGGCTGCTCCCGCCCGCAGGAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCC 1467
DB 632 GACGGGTGGCTGCTCCCGCCCGCAGGAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCC 573
QY 1468 AACAGACAGCGCGGCCACCCACTCCCGGTGACCGCCCGCAGCCACATGGACACAGCCCT 1527
DB 572 AACAGACAGCGCGGCCACCCACTCCCGGTGACCGCCCGCAGCCACATGGACACAGCCCT 513
QY 1528 CGCCTCCCGCCCGCTTTCTGCTGCTTTCTACCGACATGTAGCCCGCAGCCGCT 1587
DB 512 CGCCTCCCGCCCGCTTTCTGCTGCTTTCTACCGACATGTAGCCCGCAGCCGCT 453
QY 1588 GCCCCACTGCTCCCGCGGAGTCTGGGAGCTTCCCTGGGGAGCGGGAGGAGGA 1647
DB 452 GCCCCANCTG-CCTCCCGGAGTACTTGGGAGCTTCCCTGGGGAGCGGGAGGAGGA 394
QY 1648 GGCAGGACTCTTTGGACAGAGGCTGGGCCCTCAGTGGAGCTGCTGCCACAGCCCTG 1707
DB 393 GGCAGGACTCTTTGGACAGAGGCTGGGCCCTCAGTGGAGCTGCTGCCACAGCCCTG 334
QY 1708 GGCAGGACTCAGAGCGGAGCGGAGAACTGAGTGAAGCCCTGCTGCTGGTCTCAGGA 1767
DB 333 GGCAGGACTCAGAGCGGAGCGGAGAACTGAGTGAAGCCCTGCTGCTGGTCTCAGGA 274
QY 1768 TGGTCTCGGGGCTCTGTTTCATCAGACACCCCTTGCACCAGCTCACACATCTTCA 1827
DB 273 TGGTCTCGGGGCTCTGTTTCATCAGACACCCCTTGCACCAGCTCACACATCTTCA 214
QY 1828 TCACAGCAACCGCAGGACTTGGTCCCGCTCCTCAGAACTCAGAGCCATTCGCTCC 1887
|||||

DB 213 TCACGACAAACCCAGGACTTGGBTCCCTCCCATCTCAGTACTCACAGCCATTGCTCCC 154
QY 1888 CAGCTGGGAACTCAACCTCCCTCCCTCGCTCGTGGTGACAGAGGGGTGGACAGGG 1947
DB 153 CRGCTGGGAACTCAACCTCCCTCCCTCGCTCGTGGTGACAGAGGGGTGGACAGGG 94
QY 1948 GCGGGGGTTCCTCCCTGTACATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2007
DB 93 GCGGGGGTTCCTCCCTGTAAATAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 34
QY 2008 TTTTGTATTTTATTTTATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2040
DB 33 TTTTGTATTTTATTTTATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1
RESULT 2
AL533351 967 bp mRNA linear EST 13-FEB-2001
LOCUS AL533351 LTI_FL015_Brn1 Homo sapiens cDNA clone CS0DN004Y122 5
DEFINITION prime, mRNA sequence.
ACCESSION AL533351
VERSION AL533351.1 GI:12796844
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 967)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr.
FEATURES
1..967
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DN004Y122"
/clone_lib="LTI_FL015_Brn1"
/sex="male"
/tissue.type="Adult brain"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 173 a 325 c 286 g 177 t 6 others
ORIGIN

Query Match 44.3%; Score 923.6; DB 9; Length 967;
Best Local Similarity 98.8%; Pred. No. 1.6e-121;
Matches 956; Conservative 5; Mismatches 4; Indels 3; Gaps 3;

QY 1083 GGAGACGGGGCTGCTCAGCGCCATCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 1142
DB 1 GGAGACGGGGCTGCTCAGCGCCATCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 60
QY 1143 GCGGACCGGGTGGACATGCTCAGGAGCGGCTGCTGAGGCGCTAAAGCTCTACGTCG 1202
DB 61 GCGGACCGGGTGGACATGCTCAGGAGCGGCTGCTGAGGCGCTAAAGCTCTACGTCG 120
QY 1203 GAAGGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGAAGATTACTGACCT 1262
DB 121 GAAGGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGAAGATTACTGACCT 180
QY 1263 GCGAAGCATCAGCGCCAGCGGGGCTGAGCGGGTGTATCATCGCTGAAGATGAGATCCGGG 1322
|||||

```
Db 181 GCGAA-CATCAGCGCCAAAGGGGCTGAGGGGTGATCAGCTGAGATGGAGATCCCGGG 239
Qy 1323 CTCATCCCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGCTTGAGACTCTGAG 1382
Db 240 CTCATCCCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGCTTGAGACTCTGAG 299
Qy 1383 CGGACAGCCGGGGTGGGGGGGAGCGGGGTGGCTGCCCGCCCGCCAGCAGCTG 1442
Db 300 CGGACAGCCGGGGTGGGGGGGAGCGGGGTGGCTGCCCGCCCGCCAGCAGCTG 359
Qy 1443 TAGCCCGAGCTCAGCCCGCCAGCTCCAAAGAGAGCAGCGCGGCCACCTCCCGGTGACC 1502
Db 360 TAGCCCGAGCTCAGCCCGCCAGCTCCAAAGAGAGCAGCGCGGCCACCTCCCGGTGACC 419
Qy 1503 GCCACGCCACATGGACACAGCCCTGCCCTCCCGCCCGGGTTCCTGCTTTTAC 1562
Db 420 GCCACGCCACATGGACACAGCCCTGCCCTCCCGCCCGGGTTCCTGCTTTTAC 479
Qy 1563 GACCATGTGACCCGACACAGCCCTGCCCTCCCGCCCGGGTTCCTGCTTTTAC 1622
Db 480 GACCATGTGACCCGACACAGCCCTGCCCTCCCGCCCGGGTTCCTGCTTTTAC 539
Qy 1623 TTCCTCTGGGGAGCGGGAGGAGGAGCGAGCTCTTGGACAGAGGCTTGGCCCTCA 1682
Db 540 TTCCTCTGGGGAGCGGGAGGAGGAGCGAGCTCTTGGACAGAGGCTTGGCCCTCA 599
Qy 1683 GTGAGTGCCTGTCTCCACAGCCTGGGCTGACGTACAGAGCCGAGGCGGAACTGAGTG 1742
Db 600 GTGAGTGCCTGTCTCCACAGCCTGGGCTGACGTACAGAGCCGAGGCGGAACTGAGTG 659
Qy 1743 AGGCCCCCTGGTCTGAGTCTCAGATGGGTCT- GGGGGCTCTGTGTTCATCAGACACC 1801
Db 660 AGGCCCCCTGGTCTGAGTCTCAGATGGGTCT- GGGGGGGCTCTGTGTTCATCAGACACC 719
Qy 1802 CCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAGCGGAGCTTGGCTCCCGCATC 1861
Db 720 CCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAGCGGAGCTTGGCTCCCGCATC 779
Qy 1862 CTCAGAACTCACAAGCATTGCTCCCGAGTGGGGAACCTCAACCTCCCGCTGCTCGG 1921
Db 780 CTCAGAACTCACAAGCATTGCTCCCGAGTGGGGAACCTCAACCTCCCGCTGCTCGG 839
Qy 1922 TTGGTGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCCCTG- TACATACCTGCCATA 1980
Db 840 TTGGTGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCCCTG- TACATACCTGCCATA 899
Qy 1981° CCAACCCAGGTATTATCTCGCTGTTTGTGTTTATTTATTTATTTTGTGTTTGTGATT 2040
Db 900 CCAACCCAGGTATTATCTCGCTGTTTGTGTTTATTTATTTATTTTGTGTTTGTGATT 959
Qy 2041 TTTTAAAT 2048
Db 960 TTTTAAAT 967

RESULT 3
AL522712/c
LOCUS AL522712 LTI_NFL004_NBC2 Homo sapiens: cDNA clone CS0DB009YG04 3
DEFINITION prime, mRNA sequence.
ACCESSION AL522712
VERSION AL522712.1 GI:12786205
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 978)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
```

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 978
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 167 a 293 c 331 g 177 t 10 others
ORIGIN

Query Match 44.3%; Score 923.4; DB 9; Length 978;
Best Local Similarity 98.6%; Pred. No. 1.7e-121;
Matches 954; Conservative 7; Mismatches 4; Indels 3; Gaps 3;

Qy 1074 GGATGATGCGGAGAGCGGGGTGCTCAGCGCCATCTGCTCATCTCGGAGAGCGGCAGGA 1133
Db 965 GGATGATGCGGAGAG- GGGGTGCTCAGCGCCATCTGCTCATCTGC- GAGACCGGCAGGA 908
Qy 1134 CCTGGAGCAGCGGACCGGGTGGACATCTCTCAGGAGCGCTGCTGGAGGCGCTTAAAGT 1193
Db 907 CCTGGAGCAGCGGACCGGGTGGACATCTCTCAGGAGCGCTGCTGGAGGCGCTTAAAGT 848
Qy 1194 CTAGTGGGGAAGCGGAGGCGCCAGCGCCACATGTTCCCAAGATGCTAATGAAGAT 1253
Db 847 CTAGTGGGGAAGCGGAGGCGCCAGCGCCACATGTTCCCAAGATGCTAATGAAGAT 788
Qy 1254 TACTGACTCGGAGCAGCATCAGCGCCAGGGGCTGAGCGGTGATCAGCTGAAGATGA 1313
Db 787 TACTGACTCGGAGCAGCATCAGCGCCAGGGGCTGAGCGGTGATCAGCTGAAGATGA 728
Qy 1314 GATCCCGGGTCTCATGCGGCTCTCATCCAGAAATGTTGAGAACTCAGAGGGGCTTGA 1373
Db 727 GATCCCGGGTCTCATGCGGCTCTCATCCAGAAATGTTGAGAACTCAGAGGGGCTTGA 668
Qy 1374 CACTCTCAGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGGCTGCGCCCGCCGCC 1433
Db 667 CACTCTCAGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGGCTGCGCCCGCCGCC 608
Qy 1434 AGGAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCCACAGAGAGCAGCGCGCCACCTC 1493
Db 607 AGGAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCCACAGAGAGCAGCGCGCCACCTC 548
Qy 1494 CCGGTGACCGCCACGACCATGACACAGCCCTCGCCCTCGCCCGCGGCTTCTCTGTC 1553
Db 547 CCGGTGACCGCCACGACCATGACACAGCCCTCGCCCTCGCCCGCGGCTTCTCTGTC 488
Qy 1554 CTTTCTACCGACCATGTGACCCCGCAGCGCTGCGCCCGCAGCTGCGCCCGCGGAGTA 1613
Db 487 CTTTCTACCGACCATGTGACCCCGCAGCGCTGCGCCCGCAGCTGCGCCCGCGGAGTA 429
Qy 1614 CTGGGAGCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1673
Db 428 CTGGGAGCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 369
Qy 1674 GGGCCCTCAGTGAGCTGCTGCTCCCGCAGCGCTGGGCTGAGTCTGAGGCGGAGGCGGAG 1733
Db 368 GGGCCCTCAGTGAGCTGCTGCTCCCGCAGCGCTGGGCTGAGTCTGAGGCGGAGGCGGAG 309

```

QY 1734 AACTGAGTGAGGCCCTGGTCTCGGTCTCAGGATGGGCTCTGGGGCCCTCGGTGTTTCATC 1793
Db 308 AACTGAGTGAGGCCCTGGTCTCGGTCTCAGGATGGGCTCTGGGGCCCTCGGTGTTTCATC 249
QY 1794 AAGACACCCCTGTGCCAGCTCACACATCTTCATCACCAGCAAAAGCCAGGACTTGGCT 1853
Db 248 AAGACACCCCTGTGCCAGCTCACACATCTTCATCACCAGCAAAAGCCAGGACTTGGCT 189
QY 1854 CCCCCATCTCAGAACTCAAGCCATTGCTCCCAAGCTGGGGAACCTCAACCTCCCCC 1913
Db 188 CCCCCATCTCAGAACTCAAGCCATTGCTCCCAAGCTGGGGAACCTCAACCTCCCCC 129
QY 1914 TGCCTCGGTTGTGACAGAGGGGTGGGACAGGGGGGGTTCCTCCCTGTACATACCC 1973
Db 128 TSCCTCGGTTGTGACAGAGGGGTGGGACAGGGGGGGTTCCTCCCTGTACATACCC 69
QY 1974 TCCCATACCAAGCCAGGTATTAATTCCTCGCTGGTCTTTTATTTATTTATTTTGT 2033
Db 68 TCCCATACCAAGCCAGGTATTAATTCCTCGCTGGTCTTTTATTTATTTATTTTGT 9
QY 2034 TTTGATTT 2041
Db 8 TTTGATTT 1

RESULT 4
AL556392 992 bp mRNA linear EST 16-FEB-2001
LOCUS AL556392 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK004YG05 5
DEFINITION prime, mRNA sequence.
ACCESSION AL556392
VERSION AL556392.1 GI:12899025
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 992)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK004YG05"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 228 a 319 c 269 g 173 t 3 others
ORIGIN

Query Match 43.3%; Score 903.6; DB 9; Length 992;
Best Local Similarity 98.9%; Pred. No. 1.le-118;
Matches 928; Conservative 1; Mismatches 7; Indels 2; Gaps 2;

QY 288 CTCATTGAGACCCAGACAGTTCGTGAAGAGATAGTGCACCCCTCCCTGCCACC 347
Db 35 CACCATTGAGACCCAGACAGTTCGTGAAGAGATAGTGCACCCCTCCCTGCCACC 94

```

```

QY 348 CCCTACCCCGCATCTACAGCCTTCTGCTGTCTGTCAGACAGTCTCTCAGGCTACCA 407
Db 95 CCCTACCCCGCATCTACAGCCTTCTGCTGTCTGTCAGACAGTCTCTCAGGCTACCA 154
QY 408 CTATGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCTCCGCCGCGAGCATCCAGAGAA 467
Db 155 CTATGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCTCCGCCGCGAGCATCCAGAGAA 214
QY 468 CATGGTGTACACGCTGTACCGGGACAAAGAACTGCATCATCAAGGTGACCCGGAACCC 527
Db 215 CATGGTGTACACGCTGTACCGGGACAAAGAACTGCATCATCAAGGTGACCCGGAACCC 274
QY 528 CTCGCCAGTACTCGCCGCTGCAGAAAGTCTTTGAAGTGGGATGTCCTCAAGGAGTCTGTGAG 587
Db 275 CTCGCCAGTACTCGCCGCTGCAGAAAGTCTTTGAAGTGGGATGTCCTCAAGGAGTCTGTGAG 334
QY 588 AAACGACCCAAACAAAGAAAGAGGAGGTGCCCAAGCCGCGAGTGTCTCTGAGAGCTACAC 647
Db 335 AAACGACCCAAACAAAGAAAGAGGAGGTGCCCAAGCCGCGAGTGTCTCTGAGAGCTACAC 394
QY 648 GCTGACCGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGAAAGCGCACAGGAAACCTT 707
Db 395 GCTGACCGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGAAAGCGCACAGGAAACCTT 454
QY 708 CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAAACAGCTCAGAAACACCTGTCTC 767
Db 455 CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAAACAGCTCAGAAACACCTGTCTC 514
QY 768 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTCCACCAGTGCATCATTAAGAC 827
Db 515 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTCCACCAGTGCATCATTAAGAC 574
QY 828 TGTGGAGTTCGCAAGCAGCTGCCCGCTTCACACACCTCACCATCGCCGACAGATCAC 887
Db 575 TGTGGAGTTCGCAAGCAGCTGCCCGCTTCACACACCTCACCATCGCCGACAGATCAC 634
QY 888 CCTCCTCAAGGCTCGCTGTCGACATCTCGGATCTCGGATCTGCGACGGGTACAGCC 947
Db 635 CCTCCTCAAGGCTCGCTGTCGACATCTCGGATCTCGGATCTGCGACGGGTACAGCC 694
QY 948 CGAGCAGGACACCATGACCTTCTCGGACGGGCTGACCTTGAACCGGACCCAGATGCACAA 1007
Db 695 CGAGCAGGACACCATGACCTTCTCGGACGGGCTGACCTTGAACCGGACCCAGATGCACAA 754
QY 1008 CGTGGCTTCGGCCCCCTCACCAGCTGTGTTGCTTCGCCAACAGATGTGTCGCCCT 1067
Db 755 CGTGGCTTCGGCCCCCTCACCAGCTGTGTTGCTTCGCCAACAGATGTGTCGCCCT 814
QY 1068 GGAGATGGATGATGCGGAGACGGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGACCG 1127
Db 815 GGAGATGGATGATGCGGAGACGGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGACCG 874
QY 1128 CCAGGACCTT-GGAGCAGCGGACCGGGTGCATGCTGAGGAGCGCTGCTGTGAGGCGC 1186
Db 875 CCAGGACCTTGGAGCAGCGGACCGGGTGCATGCTGAGGAGCGCTGCTGTGAGGCGC 934
QY 1187 TAAAGTCTACGTGGGAAAGCGGAGGCCCGCGCCCC 1224
Db 935 TAAAGTCTACGTGGGAA-SGGAGGCCCGCGCCCC 971

RESULT 5
AL531149/c 931 bp mRNA linear EST 13-FEB-2001
LOCUS AL531149 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM001YN13 3
DEFINITION prime, mRNA sequence.
ACCESSION AL531149
VERSION AL531149.1 GI:12794642
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 931)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1..931
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DM001YN13"
 /clone_lib="LRI_NFL001_NBC4"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 166 a 275 c 316 g 172 t 2 others
 ORIGIN
 Query Match 42.3%; Score 882.2; DB 9; Length 931;
 Best Local Similarity 99.3%; Pred.No. 1.2e-115;
 Matches 917; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

QY 1121 GAGACCGCCAGGACCTGGAGCGCGGACC-GGGTGGACATCTGCAGGAGCGCTGCTG 1179
 DB 925 GAGACCGCCAGAGCTGGAGCGCGGACC-SGGTGGACATCTGCAGGAGCGCTGCTG 866
 QY 1180 GAGGCGCTAAGGCTACGTGGGAAGCGGAGCGCGGCCCGCCACATGTTCCCGAAG 1239
 DB 865 AAGCGCGCTAAGGCTACGTGGGAAGCGGAGCGCGGCCCGCCACATGTTCCCGAAG 806
 QY 1240 ATGCTAATGAAGATTACTGACTCGGAGGACATCAGCGCCCAAGGGGCGTACGGGGTGATC 1299
 DB 805 ATGCTAATGAAGATTACTGACTCGGAGGACATCAGCGCCCAAGGGGCGTACGGGGTGATC 746
 QY 1300 ACGTGAAGATGGAGATCCCGGGCTCCATCGCGCTCTCATCCAGGAATGTTGGAGAAC 1359
 DB 745 ACGTGAAGATGGAGATCCCGGGCTCCATCGCGCTCTCATCCAGGAATGTTGGAGAAC 686
 QY 1360 TCAGAGGCGCTGGACACTCTGAGCGGACAGCGGGGGGTGGGGGCGGAGCGGGGTGGC 1419
 DB 685 TCAGAGGCGCTGGACACTCTGAGCGGACAGCGGGGGGTGGGGGCGGAGCGGGGTGGC 626
 QY 1420 CTGCCCCCCCCCGAGGAGCTGTAGCCCCAGCCCTCAGCCCCAGCTCCCAACAGAGGACGC 1479
 DB 625 CTGCCCCCCCCCGAGGAGCTGTAGCCCCAGCCCTCAGCCCCAGCTCCCAACAGAGGACGC 566
 QY 1480 CCGGCCACCCACTCCC-CGTGACGCGCCACCGCCACATGAGACACAGCCCTCGCCCTCCGCC 1538
 DB 565 CCGGCCACCCACTCCC CGGCTGACGCCCCACGCCACATGAGACACAGCCCTCGCCCTCCGCC 506
 QY 1539 CCGGCTTTCTGCTTTCTACCGACCATGTGACCGCCGACACCGCCCTCGCCCGCCACTG 1598
 DB 505 CCGGCTTTCTGCTTTCTACCGACCATGTGACCGCCGACACCGCCCTCGCCCGCCACTG 446
 QY 1599 CCCTCCGGGAGTACTGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGGAGGAGTCTC 1658
 DB 445 CCCTCCGGGAGTACTGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGGAGGAGTCTC 386
 QY 1659 CTTGGACAGAGGCGCTGGGCGCTCAGTGAGCTGCCTGCTCCACAGCCCTGGGCTGACGTCA 1718
 DB 385 CTTGGACAGAGGCGCTGGGCGCTCAGTGAGCTGCCTGCTCCACAGCCCTGGGCTGACGTCA 326

QY 1719 GAGGCGAGGCCAGGAAGTGAAGAGGCCCTGCTGCTGGTCTCAGAGTGGGCTCTGGG 1778
 DB 325 GAGGCGAGGCCAGGAAGTGAAGAGGCCCTGCTGCTGGTCTCAGAGTGGGCTCTGGG 266
 QY 1779 GGCTCGTGTTCATCAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGC-AA 1837
 DB 265 GGCTCGTGTTCATCAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCTAA 206
 QY 1838 ACGCCAGGACTTGGCTCCCCCATCTCAGAACTCACAAGCCATTGCTCCCCAGCTGGGGA 1897
 DB 205 ACGCCAGGACTTGGCTCCCCCATCTCAGAACTCACAAGCCATTGCTCCCCAGCTGGGGA 146
 QY 1898 ACCTCAACCTCCCTCCCTCGCTCGGTGGTGACAGAGGGGTGGACAGGGGGGGGTT 1957
 DB 145 ACCTCAACCTCCCTCCCTCGCTCGGTGGTGACAGAGGGGTGGACAGGGGGGGGTT 86
 QY 1958 CCCCCTGTACATACCTGCCATACCAACCCAGCTATTATTCGCTGGTGTGTTTTT 2017
 DB 85 CCCCCTGTACATACCTGCCATACCAACCCAGCTATTATTCGCTGGTGTGTTTTT 26
 QY 2018 ATTTAATTTTTTTGTTGATT 2040
 DB 25 ATTTAATTTTTTTGTTGATT 3

RESULT 6
 BM544324
 LOCUS BM544324
 DEFINITION AGENCOURT_6490850 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5587843
 5', mRNA sequence.
 ACCESSION BM544324
 VERSION BM544324.1 GI:18775485
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1073)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Invitrogen
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12357 row: p column: 20
 High quality sequence start: 45
 High quality sequence stop: 712.
 Location/Qualifiers
 1..1073
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5587843"
 /clone_lib="NIH_MGC_125"
 /lab_host="DH10B"
 /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6; Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool of three ovaries, from females ranging in age from 38 to 49 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."

BASE COUNT 241 a 335 c 302 g 194 t 1 others
 ORIGIN

Qy	1183	CGCTAAAGTCTACGTGCGGAAGCGGAGGCCAGCGCCCCACATGTT-CCCCAAGAT	1241
Db	861	CGCTAAATCTACGTGCGGAAGCGGAGGCCAGCGCCCCACATGTTCCCCAAGAT	802
Qy	1242	GCTAATGAAGATTACTACCTCGAAGCATCAGCGCCAAAGGGGCTGACGGGTGATCAC	1301
Db	801	GCTAATGAAGATTACTGACCTCGAAGCATCAGCGCCAAAGGGGCTGACGGGTGATCAC	742
Qy	1302	GCTGAAGATGGAGATCCCGGGTCTCATCGCGCTCTCATCCAGGAAATGTTGGAGAACTC	1361
Db	741	GCTGAAGATGGAGATCCCGGGTCTCATCGCGCTCTCATCCAGGAAATGTTGGAGAACTC	682
Qy	1362	AGAGGGCTGAGACACTGATAGCGGACAGCC-GGGGGGTGGGGGGGGGNCGGGGGTGGCC	1420
Db	681	AGAGGGCTGAGACACTGTAGCGGACAGCCGGGGGGTGGGGGGGGGACGGGGGTGGCC	622
Qy	1421	TGCCCCCCCCCGCAGCAGAGTGATGCCCCAGCGCTCAGCCCCAGCTCCAACGAAGCAGCC	1480
Db	621	TGCCCCCCCCCGCAGCAGCTGTAGCCCCAGCGCTCAGCCCCAGCTCCAACGAAGCAGCC	562
Qy	1481	CGCCACCCACTCCCGTGCACCGCCAGCCACATGGACACAGCCCTCGCCCTCGCCCC	1540
Db	561	CGCCACCCACTCCCGTGCACCGCCAGCCACATGGACACAGCCCTCGCCCTCGCCCC	502
Qy	1541	GGCTTTTCTGCTTTCTACCGACCATGTGACCCCGCACCAGCCCTCGCCCC-ACCTGC	1599
Db	501	GGCTTTTCTGCTTTCTACCGACCATGTGACCCCGCACCAGCCCTCGCCCCAANCTGC	442
Qy	1600	CTCCCGGCGAGTACTGGGACCTTTCCTGGGGGACGGGGAGGAGGAGGAGGACTCC	1659
Db	441	CTCCCGGCGAGTACTGGGAGACTTTCCTGGGGGACGGGGAGGAGGAGGAGGAGGACTCC	382
Qy	1660	TTGGACAGAGGCTGGGGCCCTCAGTGGAGTGCTGCTCCCAAGCCCTGGGCTGACGTGAC	1719

1600	CTTCCCGGCGAGTACTTGGGACCTTTCCTGGGGGACGGGAGGAGGAGCGACCTCC	1659
QY		
441	CCTCCCGGCGATGACTGGGACCTTTCCTGGGGGACGGGAGGAGGAGCGACCTCC	382
DB		
1660	TTGGACAGAGGCGCTGGGGCCCTCAGTGGAGTGCTCTCTCCACAGCGCTGGGCTGACGCTCAG	1719
QY		

```
|||||
Db 381 TTGGACAGAGGCTGGGCTCAGTGAGCTGCTGCCACAGCTGGGCTGACGTCAG 322
QY 1720 AGCCGAGGCGCAGAACTAGTAGTGAGGCGCTGCTGGGCTCAGAGTGGTCTGGG 1779
Db 321 AGCCGAGGCGCAGAACTAGTAGTGAGGCGCTGCTGGGCTCAGAGTGGTCTGGG 262
QY 1780 GCTCGTGTTTCATCAGACACACCTCTGCCAGCTCAGCAGTCTTCATCAGCAGAAAC 1839
Db 261 GCTCGTGTTTCATCAGACACACCTCTGCCAGCTCAGCAGTCTTCATCAGCAGAAAC 202
QY 1840 GCAGGACTTGGCTCCGCTCCCTCAGAACTCACAAGCCATGCTCCCGAGCTGGGGAAC 1899
Db 201 GCAGGACTTGGCTCCGCTCCCTCAGAACTCACAAGCCATGCTCCCGAGCTGGGGAAC 142
QY 1900 CTCACCTCCGCTCCGCTGGTGTGACAGAGGGGTGGGAGGGGGGGGGTTC 1959
Db 141 CTCACCTCCGCTCCGCTGGTGTGACAGAGGGGTGGGAGGGGGGGGGTTC 82
QY 1960 CCTGTACATACCTCCGCTCCGCTGGTGTGACAGAGGGGTGGGAGGGGGGGGGTTC -A 2018
Db 81 CCTGTACATACCTCCGCTCCGCTGGTGTGACAGAGGGGTGGGAGGGGGGGGGTTC 22
QY 2019 TTTTAAATTTTGTGTTGAT 2039
Db 21 TTTTAAATTTTGTGTTGAT 1

RESULT 8
BQ706025
LOCUS BQ706025 885 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8351580 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6282394
5', mRNA sequence.
ACCESSION BQ706025
VERSION BQ706025.1 GI:21844924
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nhl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Mark Watson
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2476 row: 1 column: 11
High quality sequence stop: 599.
Location/Qualifiers
1..885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6282394"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 157 a 337 c 234 g 154 t 3 others
ORIGIN
```

```
Query Match 35.08; Score 729.6; DB 14; Length 885;
Best Local Similarity 91.08; Pred. No. 4.2e-94;
Matches 785; Conservative 0; Mismatches 77; Indels 1; Gaps 1;

QY 678 GAAGGTGCGCAAGCGCAGGAAACCTTCCCTGCCCTCTGCCAGCTGGGCAATACAC 737
Db 1 GAAGGTGCGCAAGCGCAGGAAACCTTCCCTGCCCTCTGCCAGCTGGGCAATACAC 60
QY 738 TAGCAACAACAGCTCAGAAACAGTGTCTCTCTGACATTTGACCTCTGGGCAAGTTTCAG 797
Db 61 TAGCAACAACAGCTCAGAAACAGTGTCTCTCTGACATTTGACCTCTGGGCAAGTTTCAG 120
QY 798 TGAATCTTCACCAAGTGCATCATTAAGACTGTGGAGTTGCGCAAGAGCTGCCGGGTT 857
Db 121 TGAATCTTCACCAAGTGCATCATTAAGACTGTGGAGTTGCGCAAGAGCTGCCGGGTT 180
QY 858 CACCACCTTACCCTCGCGGACAGATCACCCTCTCTAAGGCTGCCTGCCGTGACATCTCT 917
Db 181 CACCACCTTACCCTCGCGGACAGATCACCCTCTCTAAGGCTGCCTGCCGTGACATCTCT 240
QY 918 GATCTCGGATCTGCACGCGGTACACGCCGAGCAGGACACCATGACCTTCTCGGACGG 977
Db 241 GATCTCGGATCTGCACGCGGTACACGCCGAGCAGGACACCATGACCTTCTCGGACGG 300
QY 978 GCTGACCTGAACCGGACCCAGATGCACAGCTGGCTTCCGCCCTTACCGACCTGGT 1037
Db 301 GCTGACCTGAACCGGACCCAGATGCACAGCTGGCTTCCGCCCTTACCGACCTGGT 360
QY 1038 CTTTGGCTTCGCCAACAGCTGCTGCCCTGGAGATGATGCGGAGAGGGGCTGCT 1097
Db 361 CTTTGGCTTCGCCAACAGCTGCTGCCCTGGAGATGATGCGGAGAGGGGCTGCT 420
QY 1098 CAGCGCATCTGCTCATCTGCGGAGACCGCAGGACCTGAGAGCGCCGCGGTTGA 1157
Db 421 CAGCGCATCTGCTCATCTGCGGAGACCGCAGGACCTGAGAGCGCCGCGGTTGA 480
QY 1158 CATGCTCAGAGAGCGCTGCTGGAGCGCTAAAGGTCTACGTGGGAAGCGGAGGCCAG 1217
Db 481 CATGCTCAGAGAGCGCTGCTGGAGCGCTAAAGGTCTACGTGGGAAGCGGAGGCCAG 540
QY 1218 CCGCCCGACATGTTCCCGCAAGATGCTAAAGATTAATGACCTGCGCAAGCAGCAGCGC 1277
Db 541 CCGCCCGACATGTTCCCGCAAGATGCTAAAGATTAATGACCTGCGCAAGCAGCAGCGC 600
QY 1278 CAAGGGGCTGAGCGGTGATCAGCTGAAGATGAGATCCGGGCTCCATGCCGCTCT 1337
Db 601 CAAGGGGCTGAGCGGTGATCAGCTGAAGATGAGATCCGGGCTCCATGCCGCTCT 660
QY 1338 CATCCAGGAATGTT-GGAGAACTCAGAGGCTTGGACACTCTGAGCGGACAGCGGGGG 1396
Db 661 CATCCAGGAATGTTGGGAGAACTCAGAGGCTTGGACACTCTGAGCGGACAGCGGGGG 720
QY 1397 GTGGGGGGGGGAGGGGGTGGCTGCCCGCCCGCAGGAGCTGTAGCCCGACGCTCA 1456
Db 721 GTGGGGGGGGGAGGGGGTGGCTGCCCGCCCGCAGGAGCTGTAGCCCGACGCTCA 780
QY 1457 GCGCCAGCTCCAAAGAGAGCGCGCCGACCCACTCCCGTGCAGCGCCCGACGACATG 1516
Db 781 GCGCCAGCTCCAAAGAGAGCGCGCCGACCCACTCCCGTGCAGCGCCCGACGACATG 840
QY 1517 GACACAGCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 1539
Db 841 GCGCGTCCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 863

RESULT 9
BC030234
LOCUS BC030234
DEFINITION Homo sapiens, retinoic acid receptor, beta, clone IMAGE:5229160,
mRNA.
ACCESSION BC030234
VERSION BC030234.1 GI:20988815
KEYWORDS HTC.
```


JOURNAL COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: WGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11411 row: 1 column: 23
High quality sequence stop: 622.

FEATURES

Location/Qualifiers
1. 702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5165734"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."
BASE COUNT 145 a 237 c 212 g 108 t
ORIGIN

Query Match 31.6%; Score 659.8; DB 13; Length 702;
Best Local Similarity 98.0%; Pred. No. 3.6e-84;
Matches 689; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

QY	794	TCAGTGAACCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCG	853
Db	1	TCAGTGAACCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCG	60
QY	854	GCCTACACCCCTCACCATCGCGAGCAGATCACCTCTCCCTCAAGCTGCCTGCCTGCACA	913
Db	61	GCCTACACCCCTCACCATCGCGAGCAGATCACCTCTCCCTCAAGCTGCCTGCCTGCACA	120
QY	914	TCCTGATCTCTCGGATCTGCACGCGGTACACGCGGAGCAGGACACCATGACCTTCCTCGG	973
Db	121	TCCTGATCTCTCGGATCTGCACGCGGTACACGCGGAGCAGGACACCATGACCTTCCTCGG	180
QY	974	ACGGGCTGACCTTGAACCGGACCCAGATGACACACGCTTGGCCCTCCCTCACCGACC	1033
Db	181	ACGGGCTGACCTTGAACCGGACCCAGATGACACACGCTTGGCCCTCCCTCACCGACC	240
QY	1034	TGGTCTTTCCTTCGCGCAACCACTGCTGCCCTCGGAGATGGATGATGCGGAGACGGGG	1093
Db	241	TGGTCTTTCCTTCGCGCAACCACTGCTGCCCTCGGAGATGGATGATGCGGAGACGGGG	300
QY	1094	TGCTACGCGCATCTGCTCATCTCGGGAGACCGCAGGACCTGGAGCAGCGGACCGGG	1153
Db	301	TGCTACGCGCATCTGCTCATCTCGGGAGACCGCAGGACCTGGAGCAGCGGACCGGG	360
QY	1154	TGACATGCTGACGAGGCGCTGCTGAGGCGCTAAAGTCTACGTGCGGAAGCGGAGGC	1213
Db	361	TGACATGCTGACGAGGCGCTGCTGAGGCGCTAAAGTCTACGTGCGGAAGCGGAGGC	420
QY	1214	CCAGCGCCCCCATATGTTCCCCCAAGATGCTATGAAGATTAAGTACCTCGGAAGCATCA	1273
Db	421	CCAGCGCCCCCATATGTTCCCCCAAGATGCTATGAAGATTAAGTACCTCGGAAGCATCA	480
QY	1274	GGCCCAAGGGGCTGACGGGTGATCACGCTGAAGATGGAGATCCCGGGCTCCATGCCCG	1333
Db	481	GGCCCAAGGGGCTGACGGGTGATCACGCTGAAGATGGAGATCCCGGGCTCCATGCCCG	540
QY	1334	CTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCGCTGGACACTCTCAGCGGACACCGG	1393

Db	541	CTCTCATCCAGGAAATGTTGGAGAACTCAGAGGGCTGGACACTCTGAGCGGACGCCG	600
QY	1394	GGGTTGGGGGGCGGACGGGTGGCTGCCCGCCCGCCGCGCAGCAGCT-CTAGCCCCCAGC	1452
Db	601	GGGTTGGGGGGCGGACGGGTGGCTGCCCGCCCGCCGCGCAGCAGCTAGTAGCCCCCAGG	660
QY	1453	CTCAGCCCCAGCTCCCAACAGAGCAGCGCGGCCACCCACTCC 1495	
Db	661	CTCAGCCCCAAGATCC-ACAGAAGCAGCGCGGCAACACCTACCC 702	

RESULT 11

BQ044936/c
LOCUS BQ044936
DEFINITION UI-H-EU0-azo-e-05-0-UI-s1 NCI_CGAP_Carl Homo sapiens cDNA clone
IMAGE:5851372 3', mRNA sequence.
ACCESSION BQ044936
VERSION BQ044936.1 GI:19795852
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 654)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

FEATURES

Location/Qualifiers
1. 654
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5851372"
/clone_lib="NCI_CGAP_Carl"
/tissue_type="Osteoarthritic Cartilage"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Knee; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI_CGAP_Carl is a cDNA library containing the following tissue(s): Osteoarthritic Cartilage. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dfr)18 tail. The sequence tag for this library is TAGTACAGCT.
TAG_LIB=UI-H-EU0
TAG_TISSUE=osteothritic cartilage
TAG_SEQ=TCATCACGCT"

BASE COUNT

135 a 161 c 227 g 131 t
ORIGIN

Query Match 31.4%; Score 654; DB 14; Length 654;
Best Local Similarity 100.0%; Pred. No. 2.5e-83;
Matches 654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

```

RESULT 13
AI806984/c
LOCUS      AI806984      703 bp      mRNA      linear      EST 18-DEC-1999
DEFINITION w24908.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2356574 3' similar to gb:M73779 RETINOIC ACID RECEPTOR
ALPHA-1 (HUMAN); mRNA sequence.
ACCESSION  AI806984
VERSION     AI806984.1 GI:5393550
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 703)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 794 Std Error: 0.00.
            Seq primer: -400p from Gibco
            High quality sequence stop: 469.
FEATURES   Location/Qualifiers
            1..703
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2356574"
            /clone_lib="Soares_NFL_T_GBC_S1"
            /lab_host="DH10B"
            /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
            a modified polylinker; Site_1: Not I; Site_2: Eco RI;
            Equal amounts of plasmid DNA from three normalized
            libraries (fetal lung NBHL19W, testis NHT, and B-cell
            NCI CGAP GCBI) were mixed, and ss circles were made in
            vitro. Following HAP purification, this DNA was used as
            tracer in a subtractive hybridization reaction. The driver
            was PCR-amplified cDNAs from pools of 5,000 clones made
            from the same 3 libraries. The pools consisted of
            I.M.A.G.E. clones 297480-302087, 682632-687239,
            726408-728711, and 729096-731399. Subtraction by Bento
            Soares and M. Fatima Bonaldo."
BASE COUNT 144 a 184 c 240 g 133 t 2 others
ORIGIN
Query Match 30.9%; Score 645.4; DB 9; Length 703;
Best Local Similarity 96.5%; Pred. No. 3.8e-82;
Matches 680; Conservative 0; Mismatches 23; Indels 2; Gaps 2;
QY 1372 GACACTGTAGCGGACAGCGGGGGTGGGGGGGGGACGCGGGGTGGCGCTGCCCGCCCGCCG 1431
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
703 GACATTCTGTAGCGAAGCAGCAGAGAGT-GAGAGCGAGAGCGTAGTGG-CTGGGCGCCCGCC 646
QY 1432 CCAGGACAGCTGTAGCCCGACGCTCAGCCCGAGCTCCACAGAGAGCGCGCCGACCCAC 1491
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
645 CAGGGCAGCTGTAGCCCGACGCTCAGCCCGAGCTCCACAGAGAGCGCGCCGACCCAC 586
QY 1492 TCCCGGTGACCGCCCGACATGAGACAGCGCTCGCCCTCCCGCCCGCTTTTCT 1551
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
585 TCNCGGTGACCGCCCGACATGAGACAGCGCTCGCCCTCCCGCCCGCTTTTCT 526
QY 1552 GCCTTTCTACCGACATGTGACCGCCCGACAGCGCTCGCCCGACCTGCCCTCCCGGGCAG 1611
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
525 GCCTTTCTACCGACATGTGACCGCCCGACAGCGCTCGCCCGACCTGCCCTCCCGGGCAG 466
QY 1612 TACTGGGGACCTTCCCTGGGGGACGGGAGGAGGAGCGGAGCGGACTCCTTGGACAGAGGC 1671
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
465 TACTGGGGACCTTCCCTGGGGGACGGGAGGAGGAGGAGGAGCGGAGCGGACTCCTTGGACAGAGGC 406
QY 1672 CTGGGCGCTCAGTGGACTGCTGCTCCACAGCGCTGGCTGACCTCAGAGCGCGGCGCA 1731
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 405 CTGGGCCCTCAGTGAGCTGCCTGCTCCACAGCGCTGGGCTGAGCTCAGAGCGCGAGGCCA 346
QY 1732 GGAACGTGAGTGAGGCCCTGCTGGTCTCAGGATGGGTCTCTGGGGGCTCGTGTTC 1791
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
345 GGAACGTGAGTGAGGCCCTGCTGGTCTCAGGATGGGTCTCTGGGGGCTCGTGTTC 286
QY 1792 TCAAGACACCCCTGCGCCAGCTCCACCATCTTCATCACCAGCAACCCAGGACTGG 1851
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
285 TCAAGACACCCCTGCGCCAGCTCCACCATCTTCATCACCAGCAACCCAGGACTGG 226
QY 1852 CTCGCCCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 1911
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
225 CTCGCCCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 166
QY 1912 CTGCTCGTGGTGGTGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTATCAT 1971
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
165 CTGCTCGTGGTGGTGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTATCAT 106
QY 1972 CTGCTCATACCAACCCAGGATTAATTCCTCGCTGTTTTTTTATTTTATTTTATTTT 2031
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
105 CTGCTCATACCAACCCAGGATTAATTCCTCGCTGTTTTTTTATTTTATTTTATTTT 46
QY 2032 GTTTTGATTTTTTTTAAAGAAATTTTCATTTTAAGCACACAAAAA 2076
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45 GTTCTGATTTTTTTTAAAGAAATTTTCATTTTAAGCACACATTTAA 1
RESULT 14
LOCUS      BG028346      1164 bp      mRNA      linear      EST 24-JAN-2001
DEFINITION 602295596F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4390510 5',
            mRNA sequence.
ACCESSION  BG028346
VERSION     BG028346.1 GI:12417440
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1164)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10079 row: 0 column: 23
            High quality sequence stop: 677.
FEATURES   Location/Qualifiers
            1..1164
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4390510"
            /clone_lib="NIH_MGC_86"
            /tissue_type="osteosarcoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: bone; Vector: pCMV-Sport6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.533 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
BASE COUNT 285 a 376 c 339 g 164 t
ORIGIN
Query Match 30.6%; Score 639; DB 12; Length 1164;
Best Local Similarity 97.4%; Pred. No. 2.2e-81;
Matches 681; Conservative 0; Mismatches 15; Indels 3; Gaps 3;

```

QY 1330 CCGCCTCTCATCCAGAAATGTTGGAACTCAGAGGCGCTGGACACTCTGAGCGGACAG 1389
Db 1 CCGCCTCTCATCCAGAAATGTTGGAACTCAGAGGCGCTGGACACTCTGAGCGGACAG 60
QY 1390 CCGGGGGTGGGGGGGAGCGGGGTGGCTGCCCGCCCGCCCGCAGGACGCTGAGCCCC 1449
Db 61 CCGGGGGTGGGGGGGAGCGGGGTGGCTGCCCGCCCGCCCGCAGGACGCTGAGCCCC 120
QY 1450 AGCCTCAGCCCCAGCTCCCAACAGCAGCCCGGCCACCCACTCCCGGTACCGCCCGCAG 1509
Db 121 AGCCTCAGCCCCAGCTCCCAACAGCAGCCCGGCCACCCACTCCCGGTACCGCCCGCAG 180
QY 1510 CCACATGAGACAGCCCTCGCCCTCGCCCGCCCGCTTTCTCTGCTTTCTTACCGACCATG 1569
Db 181 CCACATGAGACAGCCCTCGCCCTCGCCCGCCCGCTTTCTCTGCTTTCTTACCGACCATG 240
QY 1570 TGACCCCGACAGCCCTCGCCCGCCCGCTCGCCCGCGGAGTACGAGGACCTTCCCTG 1629
Db 241 TGACCCCGACAGCCCTCGCCCGCCCGCTCGCCCGCGGAGTACGAGGACCTTCCCTG 300
QY 1630 GGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1689
Db 301 GGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
QY 1690 GCCTGCTCCACAGCCTGGGCTGACGCTCAGAGCGCCGAGCGGAGGAGGAGGAGGAGGAGG 1749
Db 361 GCCTGCTCCACAGCCTGGGCTGACGCTCAGAGCGCCGAGCGGAGGAGGAGGAGGAGGAGG 420
QY 1750 TGGTCTGGGCTCAGGATGGGTCTCTGGGGGCTCGTGTTCATCAAGACACCCCTCTGCC 1809
Db 421 TGGTCTGGGCTCAGGATGGGTCTCTGGGGGCTCGTGTTCATCAAGACACCCCTCTGCC 480
QY 1810 CAGCTCACCACATCTTCATCACCAGCAAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1869
Db 481 CAGCTCACCACATCTTCATCACCAGCAAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540
QY 1870 TCACAGGACATTCGCTCCCGAGTGGGAGACCTCAACCTCCCGCTGCC-TCGGTGGTGA 1928
Db 541 TCACAGGACATTCGCTCCCGAGTGGGAGACCTCAACCTCCCGCTGCCATCGGTAGGTGA 600
QY 1929 CAGAGGGGTGGGACAGGGG-CGGGGGGTTCGCCCTGTATACATACCCCTG-CCATACCAAC 1986
Db 601 CAGAGGGGTGGGACAGGGGCGGGGGTTCGCCCTGTATACATACCCCTGTCATACCAAC 660
QY 1987 CCAGTATTAATTCGCTGGGTTTCTTTTATTTAAT 2025
Db 661 CCAGTATTAATTCGCTGGTGGTTCGCTGATTTCAAT 699

RESULT 15
BM790595
LOCUS
DEFINITION K-EST0070459 S21SN520 Homo sapiens cDNA clone S21SN520-11-G11 5',
mRNA sequence.
ACCESSION BM790595
VERSION BM790595.1 GI:19138827
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 646)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel.: +82-42-860-4470

Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 11 row: 6 column: 11
High quality sequence stop: 646.
Location/Qualifiers
1. 646
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S21SN520-11-G11"
/clone_lib="S21SN520"
/sex="F"
/tissue_type="Stomach"
/cell_type="Floating aggregates"
/cell_line="SNU-520"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
BASE COUNT 140 a 213 c 184 g 109 t
ORIGIN
Query Match 30.4%; Score 635; DB 14; Length 646;
Best Local Similarity 99.8%; Pred. No. 1.2e-80;
Matches 646; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 654 GCCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGGAAAGCGCACGAGAACTTCCTCTGC 713
Db 1 GCCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGGAAAGCGCACGAGAACTTCCTCTGC 60
QY 714 CTTCTGCAGCTGGGCAATACACTACGAAACAGCTCAGAAACAGCTGCTCTCTGGA 773
Db 61 CTTCTG-CAGCTGGGCAATACACTACGAAACAGCTCAGAAACAGCTGCTCTCTGGA 119
QY 774 CATTGACCTCTGGGCAAGTTTCACTGAACCTCTCCACCAAGTGCATCTTAAGACTGTGGA 833
Db 120 CATTGACCTCTGGGCAAGTTTCACTGAACCTCTCCACCAAGTGCATCTTAAGACTGTGGA 179
QY 834 GTTCGCCAAGCAGCTGCCGGCTTACCCACCTCACCATCGCCGACGAGTACCCCTCT 893
Db 180 GTTCGCCAAGCAGCTGCCGGCTTACCCACCTCACCATCGCCGACGAGTACCCCTCT 239
QY 894 CRAAGGTGCTGCTGGACATCTCTGATCTTCCGGATCTGACGGGTACACGCCCGAGCA 953
Db 240 CRAAGGTGCTGCTGGACATCTCTGATCTTCCGGATCTGACGGGTACACGCCCGAGCA 299
QY 954 GGACACCATGACCTTCTCGGAGCGGCTGACCTTGAACCGGACCCAGATGCACACGCTGG 1013
Db 300 GGACACCATGACCTTCTCGGAGCGGCTGACCTTGAACCGGACCCAGATGCACACGCTGG 359
QY 1014 CTTTCGCCCTTACCGACCTGGTCTTTTGGCTTTCGCCAACACAGCTGCTGCCCTTGGAGAT 1073
Db 360 CTTTCGCCCTTACCGACCTGGTCTTTTGGCTTTCGCCAACACAGCTGCTGCCCTTGGAGAT 419
QY 1074 GGATGATGGGAGACGGGGCTGCTCAGGCCCATCTGCCTCATCTGCGGAGACGCCAGCA 1133
Db 420 GGATGATGGGAGACGGGGCTGCTCAGGCCCATCTGCCTCATCTGCGGAGACGCCAGCA 479
QY 1134 CTTGGAGAGCGGACCGGCTGGACATGCTGAGGAGCGGCTGCTGAGGAGCGCTAAAGGT 1193
Db 480 CTTGGAGAGCGGACCGGCTGGACATGCTGAGGAGCGGCTGCTGAGGAGCGCTAAAGGT 539

QY 1194 CTACGTGCGGAAGCGAGCGCCAGCGCCGCCCCACATGTTCCCCAAGATGCTAATGAAGAT 1253
 |||||
 Db 540 CTACGTGCGGAAGCGAGCGCCAGCGCCGCCCCACATGTTCCCCAAGATGCTAATGAAGAT 599
 |||||
 QY 1254 TACTGACCTGCGAAGCATCAGCGCCCAAGGGGGCTGAGCGGGTGATCA 1300
 |||||
 Db 600 TACTGACCTGCGAAGCATCAGCGCCCAAGGGGGCTGAGCGGGTGATCA 646
 |||||

Search completed: March 29, 2003, 15:17:39
 Job time : 2342 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 09:49:54 ; Search time 70 Seconds
(without alignments)
9138.968 Million cell updates/sec

Title: US-09-691-220-1
Perfect score: 2086
Sequence: 1 aacagcacgagcgagggg.....cacaaaaa.....2086

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1782.2	85.4	3036	1	US-08-306-691B-52
2	1782.2	85.4	3036	2	US-08-095-728B-1
3	1782.2	85.4	3036	5	PCT-US92-02320A-1
4	1782.2	85.4	3511	3	US-08-892-747-13
5	1778.8	85.3	2928	2	US-08-095-728B-3
6	1778.8	85.3	2928	5	PCT-US92-02320A-3
7	1777.2	85.2	2940	2	US-08-592-383-1
8	1777.2	85.2	2940	6	5171671-1
9	1200	57.5	2658	2	US-08-592-383-3
10	749.2	35.9	1576	6	5260432-1
11	702.4	33.7	704	2	US-08-592-383-5
12	667.8	32.0	2989	6	5223606-1
13	293.4	14.1	558	2	US-08-896-365-5
14	199.4	9.6	201	2	US-08-592-383-6
15	183	8.8	1866	1	US-08-336-408B-1
16	183	8.8	1866	5	PCT-US91-00399-1
17	180.2	8.6	1893	6	5438126-1
18	177.6	8.5	2095	1	US-08-336-408B-3
19	177.6	8.5	2095	5	PCT-US91-00399-3
20	160.8	7.7	2204	1	US-07-952-800-3
21	160.8	7.7	2219	1	US-08-336-408B-7
22	160.8	7.7	2219	5	PCT-US91-00399-7
23	159.6	7.7	816	1	US-08-485-971-22
24	159.6	7.7	816	1	US-08-383-754-22
25	159.6	7.7	816	1	US-08-485-978-22
26	159.6	7.7	816	2	US-08-486-814-22
27	159.6	7.7	816	2	US-08-487-472-22

28	159.6	7.7	816	3	US-08-485-740-22	Sequence 22, Appl
29	159.6	7.7	816	3	US-09-162-184-22	Sequence 22, Appl
30	159.6	7.7	816	4	US-09-161-902-22	Sequence 22, Appl
31	159.6	7.7	816	4	US-09-489-777A-22	Sequence 22, Appl
32	159.6	7.7	816	5	PCT-US95-08179-22	Sequence 22, Appl
33	159.6	7.7	1443	3	US-09-162-184-37	Sequence 37, Appl
34	159.6	7.7	1443	4	US-09-489-777A-37	Sequence 37, Appl
35	157	7.5	1688	2	US-08-649-619B-2	Sequence 2, Appl
36	157	7.5	1898	1	US-08-342-411A-1	Sequence 1, Appl
37	157	7.5	1979	2	US-08-649-619B-1	Sequence 1, Appl
38	157	7.5	2030	1	US-08-330-518-1	Sequence 1, Appl
39	157	7.5	2030	2	US-08-330-283-1	Sequence 1, Appl
40	157	7.5	2030	2	US-08-646-248-1	Sequence 1, Appl
41	157	7.5	2030	5	PCT-US95-13924-1	Sequence 1, Appl
42	157	7.5	2030	5	PCT-US95-13931-1	Sequence 1, Appl
43	155.4	7.4	1813	5	PCT-US94-12883-3	Sequence 3, Appl
44	150.8	7.2	1934	4	US-08-776-844-1	Sequence 1, Appl
45	150.8	7.2	1959	1	US-08-342-411A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-306-691B-52
; Sequence 52, Application US/08306691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Conda, Lavoragna & Monaco, P.C.
; STREET: Two Penn Center, Suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306.691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: No. 5734039e
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3036 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-306-691B-52

Query Match 85.4%; Score 1782.2; DB 1; Length 3036;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 290 CCATTGAGACCCAGCAGCAGTTCGTGAGAGAGTAGTCCCGCCCTCCCTCGCCACCCC 349

QY 2030 TTGTTTGGATTTTAAATAAGAAATTTTCATTTTAAAGCACAAATAA 2076
|||||
Db 2990 TTGTTTGGATTTTAAATAAGAAATTTTCATTTTAAAGCACAAATAA 3036
|||||

RESULT 3

PCT-US92-02320A-1
; Sequence 1, Application PC/TUS9202320A
; GENERAL INFORMATION:
; APPLICANT: Sloan-Kettering Institute, For Cancer Research
; TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM
; STREET: 30 ROCKEFELLER PLAZA
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/02320A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 675,084
; FILING DATE: 22-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 673,838
; FILING DATE: 22-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WHITE, JOHN P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 38694-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 644-0525
; TELEX: (212) 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3036 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: MYL-RAR
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 67..2457
; OTHER INFORMATION:
PCT-US92-02320A-1

Query Match 85.4%; Score 1782.2; DB 5; Length 3036;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 290 CCATTGAGACCCAGCAGCAGGTTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACCCC 349
|||||
Db 1250 CCATTGAGACCCAGCAGCAGGTTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACCCC 1309
|||||
QY 350 CTCTACCCCGCATCTACAAGCCTTGCTTGTCTCTCAGGACAAGTCTCTCAGGCTACCCT 409
|||||
Db 1310 CTCTACCCCGCATCTACAAGCCTTGCTTGTCTCTCAGGACAAGTCTCTCAGGCTACCCT 1369
|||||
QY 410 ATGGGGTACGCGCTGTGAGGGTGTCAAGGGCTTCTTCCGCGCAGCATCCAGAAGAACA 469
|||||
Db 1370 ATGGGGTACGCGCTGTGAGGGTGTCAAGGGCTTCTTCCGCGCAGCATCCAGAAGAACA 1429
|||||

QY 470 TGTGTACAGCTGTACCCGGGCAAGAAGTGTGATCATCAAGAAGTGACCCGGAAACCCCT 529
|||||
Db 1430 TGTGTACAGCTGTACCCGGGCAAGAAGTGTGATCATCAAGAAGTGACCCGGAAACCCCT 1489
|||||
QY 530 GCCAGTACTGCCGACTGTCAGAAAGTGTCTTGAAGTGGGCATGTCCAAAGAGTCTGTGAGAA 589
|||||
Db 1490 GCCAGTACTGCCGACTGTCAGAAAGTGTCTTGAAGTGGGCATGTCCAAAGAGTCTGTGAGAA 1549
|||||
QY 590 ACAGCCGAAACAAGAAGAAGAGAGTGTCCCAAGCCCGAGTGTCTCTGAGAGCTACACGC 649
|||||
Db 1550 ACAGCCGAAACAAGAAGAAGAGAGTGTCCCAAGCCCGAGTGTCTCTGAGAGCTACACGC 1609
|||||
QY 650 TGACGCCGAGGTGGGGAGTCTATTGAGAAAGTGGCGCAAGCGCACAGAAACCTTCC 709
|||||
Db 1610 TGACGCCGAGGTGGGGAGTCTATTGAGAAAGTGGCGCAAGCGCACAGAAACCTTCC 1669
|||||
QY 710 CTGCCCTCTGCCAGCTGGGCAATACACTACGAAACACAGCTCAGAACACAGTGTCTCTC 769
|||||
Db 1670 CTGCCCTCTGCCAGCTGGGCAATACACTACGAAACACAGCTCAGAACACAGTGTCTCTC 1729
|||||
QY 770 TGGACATTGACCTCTGGGACAAAGTTCAGTGAAGTCTCCACCAAGTGCATCATTAAGACTG 829
|||||
Db 1730 TGGACATTGACCTCTGGGACAAAGTTCAGTGAAGTCTCCACCAAGTGCATCATTAAGACTG 1789
|||||
QY 830 TGGAGTTCGCCAAGCAGCTGCCCGGCTTCACCAACCTCACCATTGCGCGACAGATCACCC 889
|||||
Db 1790 TGGAGTTCGCCAAGCAGCTGCCCGGCTTCACCAACCTCACCATTGCGCGACAGATCACCC 1849
|||||
QY 890 TCCTCAAGCTGCTGCTGGACATCTCTGCGGATCTGCAGCGGTTAGACGCGCG 949
|||||
Db 1850 TCCTCAAGCTGCTGCTGGACATCTCTGCGGATCTGCAGCGGTTAGACGCGCG 1909
|||||
QY 950 AGCAGGACACCATGACCTTCTCGGACGGCTGACCTCTGAACCGGACCCAGATGCACAAACG 1009
|||||
Db 1910 AGCAGGACACCATGACCTTCTCGGACGGCTGACCTCTGAACCGGACCCAGATGCACAAACG 1969
|||||
QY 1010 CTGGCTTGGCGCCCTCAGCAGCTGGTCTTTGGCTTTGCGCAACCAAGCTGCTGCCCTGG 1069
|||||
Db 1970 CTGGCTTGGCGCCCTCAGCAGCTGGTCTTTGGCTTTGCGCAACCAAGCTGCTGCCCTGG 2029
|||||
QY 1070 AGATGGATGATCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGACGCGC 1129
|||||
Db 2030 AGATGGATGATCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGACGCGC 2089
|||||
QY 1130 AGGACCTGGAGCAGCGGACCGGGTGGACATGCTGAGGAGCGGCTGCTGGAGCGCGTAA 1189
|||||
Db 2090 AGGACCTGGAGCAGCGGACCGGGTGGACATGCTGAGGAGCGGCTGCTGGAGCGCGTAA 2149
|||||
QY 1190 AGGTCTAGTGGGAAAGCGGAGGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGA 1249
|||||
Db 2150 AGGTCTAGTGGGAAAGCGGAGGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGA 2209
|||||
QY 1250 AGATTACTGACCTGCGAAGCATCAGCGCAAGGGGCTGAGCGGTGATCAGCTGAAGA 1309
|||||
Db 2210 AGATTACTGACCTGCGAAGCATCAGCGCAAGGGGCTGAGCGGTGATCAGCTGAAGA 2269
|||||
QY 1310 TGGAGATCCCGGCTCCATCGCGCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC 1369
|||||
Db 2270 TGGAGATCCCGGCTCCATCGCGCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC 2329
|||||
QY 1370 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGGGAGCGGGGTGCGCTGCCCGCCC 1429
|||||
Db 2330 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGGGAGCGGGGTGCGCTGCCCGCCC 2389
|||||
QY 1430 CGCAGGAGCTGTAGCCCCAGCTCAGCCCCAGCTCCAAACAAGACGCCCGGCCACCC 1489
|||||
Db 2390 CGCAGGAGCTGTAGCCCCAGCTCAGCCCCAGCTCCAAACAAGACGCCCGGCCACCC 2449
|||||
QY 1490 ACTCCCGTGACCGCCCGACGCCCATGTGACACAGCCCTCGGCCCTCGGCCCGGGCTTTCT 1549
|||||
Db 2450 ACTCCCGTGACCGCCCGACGCCCATGTGACACAGCCCTCGGCCCTCGGCCCGGGCTTTCT 2509
|||||

Db 2445 CTGGCTTGGGCGCCCTACCGACCTGCTTGGCTTGGCAACACGAGCTGCTGCCCTGG 2504
QY 1070 AGATGGATGATCGGAGACGGGCTGCTCAGCGCATCTGCTCATCTCGGAGACGCC 1129
Db 2505 AGATGGATGATCGGAGACGGGCTGCTCAGCGCATCTGCTCATCTCGGAGACGCC 2564
QY 1130 AGGACCTGGAGACCGGACCGGCTGGACATCTGCTCAGAGCGGCTGCTGGAGCGCTAA 1189
Db 2565 AGGACCTGGAGACCGGACCGGCTGGACATCTGCTCAGAGCGGCTGCTGGAGCGCTAA 2624
QY 1190 AGGCTACGTCGGAAGCGGAGCGGCTGGAGACATCTGCTCAGAGCGGCTGCTGGAGCGCTAA 1249
Db 2625 AGGCTACGTCGGAAGCGGAGCGGCTGGAGACATCTGCTCAGAGCGGCTGCTGGAGCGCTAA 2684
QY 1250 AGATTACTGACCTCGAAGCATCAGCGGCTGAGCGGCTGAGCGGCTGATCAGCTGAAGA 1309
Db 2685 AGATTACTGACCTCGAAGCATCAGCGGCTGAGCGGCTGAGCGGCTGATCAGCTGAAGA 2744
QY 1310 TGGAGATCCCGGGCTCCATCCCGCTCTCATCCAGGAATTTTGGAGAACTCAGAGGGCC 1369
Db 2745 TGGAGATCCCGGGCTCCATCCCGCTCTCATCCAGGAATTTTGGAGAACTCAGAGGGCC 2804
QY 1370 TGGACACTCTGAGGACACCGGCGGCTGGGGGCGGAGCGGCTGCTGCCCGCCC 1429
Db 2805 TGGACACTCTGAGGACACCGGCGGCTGGGGGCGGAGCGGCTGCTGCCCGCCC 2864
QY 1430 CGCAGGACGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCT 1489
Db 2865 CGCAGGACGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCT 2924
QY 1490 ACTCCCGCTGACCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCT 1549
Db 2925 ACTCCCGCTGACCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCT 2984
QY 1550 CTGCTCTTACCGACCATGTGACCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCT 1609
Db 2985 CTGCTCTTACCGACCATGTGACCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCT 3044
QY 1610 AGTACTGGGACCTTCCCTGGGGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669
Db 3045 AGTACTGGGACCTTCCCTGGGGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3104
QY 1670 GCTGGGCGCTCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1729
Db 3105 GCTGGGCGCTCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789
QY 1730 CAGGAACCTGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3164
Db 3165 CAGGAACCTGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3224
QY 1790 CATCAGACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1849
Db 3225 CATCAGACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3284
QY 1850 GCTCCCGCATCTCAGACTCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1909
Db 3285 GCTCCCGCATCTCAGACTCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3344
QY 1910 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1969
Db 3345 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3404
QY 1970 ACCCTGCCATACCAACCGAGTATTAATCTCGTGGTGGTGGTGGTGGTGGTGGTGGT 2029
Db 3405 ACCCTGCCATACCAACCGAGTATTAATCTCGTGGTGGTGGTGGTGGTGGTGGTGGT 3464
QY 2030 TTGTTTTCATTTTTTATAGAAATTTTCATTTTAAAGCAAAAAA 2076
Db 3465 TTGTTTTCATTTTTTATAGAAATTTTTCATTTTAAAGCAAAAAA 3511

RESULT 5

US-08-095-728B-3
; Sequence 3, Application US/08095728B
; Patent No. 5843642
; GENERAL INFORMATION:
; APPLICANT: DMITROVSKY, ETHAN
; APPLICANT: WARRELL JR, RAYMOND P
; APPLICANT: MILLER JR, WILSON H
; APPLICANT: FRANKEL, STANLEY
; TITLE OF INVENTION: METHODS FOR THE DETECTION AND
; TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM LLP
; STREET: 1185 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,728B
; FILING DATE: 21-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/673,838
; FILING DATE: 22-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WHITE, JOHN P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 38694-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: hRAR ALPHA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 103..1488
; OTHER INFORMATION:
US-08-095-728B-3

Query Match 85.3%; Score 1778.8; DB 2; Length 2928;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1786; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 288 CTCATTGAGACCCAGAGCAGCAGTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC 347
Db 279 CACCATGAGACCCAGAGCAGCAGTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC 338
QY 348 CCCTCTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAGTCTCTCAGGCTACCA 407
Db 339 CCCTCTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAGTCTCTCAGGCTACCA 398
QY 408 CTATGGGGTCAGGCGCTGTGAGGGCTGCAAGGCTTCTTCCCGCGCAGCATCCAGAAGAA 467
Db 399 CTATGGGGTCAGGCGCTGTGAGGGCTGCAAGGCTTCTTCCCGCGCAGCATCCAGAAGAA 458
QY 468 CATGTTGTACAGTGTCCCGGACAGAACTGCATCATCAACAAGGTGACCCGGAACCC 527
Db 459 CATGTTGTACAGTGTCCCGGACAGAACTGCATCATCAACAAGGTGACCCGGAACCC 518
QY 528 CTGCCAGTACTCCGACTGCAGAAAGTCTTTGAAGTGGGCTATGTCACAAAGGAGTCTGTGAG 587

Db 519 CTGCCAGTACTGCGGACTCAGAAAGTCTTTGAAGTGGGATGTCACAAAGGAGTCTGTGAG 578
QY 588 AAACGACCGAACAAGAAAGAGAGTGGCCAAAGCCGAGTGTCTGTGAGAGCTACAC 647
Db 579 AAACGACCGAACAAGAAAGAGAGTGGCCAAAGCCGAGTGTCTGTGAGAGCTACAC 638
QY 648 GCTGACGCGGAGGTGGGGAGTCTATTGAGAAAGTGGCAGAAAGCCAGCCAGCAACCTT 707
Db 639 GCTGACGCGGAGGTGGGGAGTCTATTGAGAAAGTGGCAGAAAGCCAGCCAGCAACCTT 698
QY 708 CCTCGCCTCTGCCAGCTGGGCAATACACTACGAAACAGCTCAGAAACAGCTGTCTC 767
Db 699 CCTCGCCTCTGCCAGCTGGGCAATACACTACGAAACAGCTCAGAAACAGCTGTCTC 758
QY 768 TCTGACATTGACCTTGGGACAAAGTTCAAGTGAAGTCTCCACAAAGTGAATCAATTAAGAC 827
Db 759 TCTGACATTGACCTTGGGACAAAGTTCAAGTGAAGTCTCCACAAAGTGAATCAATTAAGAC 818
QY 828 TGTGAGTTGCGCAAGCAGCTCCCGGCTTACACACCTTACCATCGCCGACCCAGATCAC 887
Db 819 TGTGAGTTGCGCAAGCAGCTCCCGGCTTACACACCTTACCATCGCCGACCCAGATCAC 878
QY 888 CTTCTCAAGGCTGCTGCTGGACATCTGATCTGCGGATCTGCACGCGGTACAGCC 947
Db 879 CTTCTCAAGGCTGCTGCTGGACATCTGATCTGCGGATCTGCACGCGGTACAGCC 938
QY 948 CGAGCAGACACCATGACCTTCTCGAGGGGTGACCTGACCTGAACCGGACCCAGATGACAA 1007
Db 939 CGAGCAGACACCATGACCTTCTCGAGGGGTGACCTGACCTGAACCGGACCCAGATGACAA 998
QY 1008 CGCTGGCTTGGGCGCCCTTACCGGCTTCTGCTTTCCTTCCCTTCCCAACAGCTGCTGCCCT 1067
Db 999 CGCTGGCTTGGGCGCCCTTACCGGCTTCTGCTTTCCTTCCCTTCCCAACAGCTGCTGCCCT 1058
QY 1068 GGAGATGATGATGCGGAGAGCGGGCTGCTCAGCGCCATCTGCCCTCATCTGGGAGACCG 1127
Db 1059 GGAGATGATGATGCGGAGAGCGGGCTGCTCAGCGCCATCTGCCCTCATCTGGGAGACCG 1118
QY 1128 CCAGGACCTGAGCAGCGGACCGGGTGGACATGCTCAGGAGCGCTGCTGGAGCGCT 1187
Db 1119 CCAGGACCTGAGCAGCGGACCGGGTGGACATGCTCAGGAGCGCTGCTGGAGCGCT 1178
QY 1188 AAAGTCTACCTGCGGAAGCGGAGCGCCAGCGCGCCCTTCCCTTCCCAACAGCTGCTTAT 1247
Db 1179 AAAGTCTACCTGCGGAAGCGGAGCGCCAGCGCGCCCTTCCCTTCCCAACAGCTGCTTAT 1238
QY 1248 GAAGATTACTGACCTGCGGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGATCAGCTGAA 1307
Db 1239 GAAGATTACTGACCTGCGGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGATCAGCTGAA 1298
QY 1308 GATGAGATCCCGGCTCCATGCCCTCTCATCAGGAAATGTTGGAGAACTCAGAGG 1367
Db 1299 GATGAGATCCCGGCTCCATGCCCTCTCATCAGGAAATGTTGGAGAACTCAGAGG 1358
QY 1368 CTTGACACTGAGCGACAGCGGGGTGGGGGGGAGCGGGGTGGGCTGCCCTGCCCT 1427
Db 1359 CTTGACACTGAGCGACAGCGGGGTGGGGGGGAGCGGGGTGGGCTGCCCTGCCCT 1418
QY 1428 CCCGCCAGGAGCTGTAGCCCGCAGCTCAGCGCCAGCTTCCACAGAAAGCGCCGCCAC 1487
Db 1419 CCCGCCAGGAGCTGTAGCCCGCAGCTCAGCGCCAGCTTCCACAGAAAGCGCCGCCAC 1478
QY 1488 CCACTCCCGGTAGCCCGCCAGCCATGGACAGCGCTCGCCCTCGCCCGCGGCTTTT 1547
Db 1479 CCACTCCCGGTAGCCCGCCAGCCATGGACAGCGCTCGCCCTCGCCCGCGGCTTTT 1538
QY 1548 CTCTGCTTTTCTACGACCATGTGACCCCGACAGCGCTTGGCCCGCCAGCTGCCCTGCCG 1607
Db 1539 CTCTGCTTTTCTACGACCATGTGACCCCGACAGCTTGGCCCGCCAGCTGCCCTGCCG 1598
QY 1608 GCAGTACTGGGAGCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667

RESULT 6

PCT-US92-02320A-3

; Sequence 3, Application PC/TUS9202320A

; GENERAL INFORMATION:

; APPLICANT: Sloan-Kettering Institute, For Cancer Research

; TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: COOPER & DUNHAM

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: U.S.A.

; ZIP: 10112

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.24

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/02320A

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 675,084

; FILING DATE: 22-MAR-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 673,838

; FILING DATE: 22-MAR-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: WHITE, JOHN P

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 38694-PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 977-9550

; TELEFAX: (212) 644-0525

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2928 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: hRAR ALPHA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 103..1488
; OTHER INFORMATION:
PCT-US92-02320A-3

Query Match 85.3%; Score 1778.8; DB 5; Length 2928;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1786; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 288 CTCATTGAGACCCAGACAGCAGCTCTGGAAGAGATAGTCCAGAGCCTCCCTCGCCACC 347
|
Db 279 CACCATGACACCCAGACAGCAGCTCTGGAAGAGATAGTCCAGAGCCTCCCTCGCCACC 338
|
QY 348 CCCTCTACCCCGCATCTACAAAGCCTTCTGTTGTCTGTCAGACAAAGTCTCAGGCTACCA 407
|
Db 339 CCCTCTACCCCGCATCTACAAAGCCTTCTGTTGTCTGTCAGACAAAGTCTCAGGCTACCA 398
|
QY 408 CTATGGGGTCAGCGCCTGTGAGGGCTCCAAAGGGCTTCTCCGCGCAGCATCCAGAGAA 467
|
Db 399 CTATGGGGTCAGCGCCTGTGAGGGCTCCAAAGGGCTTCTCCGCGCAGCATCCAGAGAA 458
|
QY 468 CATGGTGTACACGTGTACCCGGGACAGAACTGCATCATCAACAAAGTGTACCCGGAAACC 527
|
Db 459 CATGGTGTACACGTGTACCCGGGACAGAACTGCATCATCAACAAAGTGTACCCGGAAACC 518
|
QY 528 CTCGCCAGTACTGCCAGTGCAGAGTCTTGAAGTGGGATGTCCAAAGGAGTCTGTGAG 587
|
Db 519 CTCGCCAGTACTGCCAGTGCAGAGTCTTGAAGTGGGATGTCCAAAGGAGTCTGTGAG 578
|
QY 588 AACGACCCGAAACAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTCAGAGCTACAC 647
|
Db 579 AACGACCCGAAACAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTCAGAGCTACAC 638
|
QY 648 GCTGACCCGCGAGTGGGGAGCTCATTTGAGAGAGTGGCGAAAGCGCACAGAAACCTT 707
|
Db 639 GCTGACCCGCGAGTGGGGAGCTCATTTGAGAGAGTGGCGAAAGCGCACAGAAACCTT 698
|
QY 708 CCTGCCCTCTGCCAGTGGGCAAAATACACTAGCAACAAAGCTCAGAAACAGTGTCTC 767
|
Db 699 CCTGCCCTCTGCCAGTGGGCAAAATACACTAGCAACAAAGCTCAGAAACAGTGTCTC 758
|
QY 768 TCTGGACATGACCTCTGGGACAAAGTTCAGTGAATCTCCACCAAGTGCATATTAGAC 827
|
Db 759 TCTGGACATGACCTCTGGGACAAAGTTCAGTGAATCTCCACCAAGTGCATATTAGAC 818
|
QY 828 TGTGAGTTTCGCGCAAGCAGCTGCCGGCTTCACCAACCTCACCATGCCGACCAAGTAC 887
|
Db 819 TGTGAGTTTCGCGCAAGCAGCTGCCGGCTTCACCAACCTCACCATGCCGACCAAGTAC 878
|
QY 888 CTTCTCAAGGTTGCTGCTGGACATCTCTGATCTCTGCGGATCTGACGCGGTACACGCC 947
|
Db 879 CTTCTCAAGGTTGCTGCTGGACATCTCTGATCTCTGCGGATCTGACGCGGTACACGCC 938
|
QY 948 CGAGCAGGACACCATGACCTTCTCGACCGGGTGCACCTGAAACCGGACCCAGATGACAA 1007
|
Db 939 CGAGCAGGACACCATGACCTTCTCGACCGGGTGCACCTGAAACCGGACCCAGATGACAA 998
|
QY 1008 CGCTGGCTTCGGCCCTCACCAGACCTGTCTTTGCTTCGCCAACCAGCTGCGCCCT 1067
|
Db 999 CGCTGGCTTCGGCCCTCACCAGACCTGTCTTTGCTTCGCCAACCAGCTGCGCCCT 1058
|
QY 1068 GGAGATGGATGATGGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGACCG 1127
|
Db 1059 GGAGATGGATGATGGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGACCG 1118
|
QY 1128 CCAGGACCTGGAGACCGGACCGGCTGGACATGTCTGAGGAGCCGCTGCTGAGGCGCT 1187
|
Db 1119 CCAGGACCTGGAGACCGGACCGGCTGGACATGTCTGAGGAGCCGCTGCTGAGGCGCT 1178
|

RESULT 7

US-08-592-383-1

; Sequence 1, Application US/08592383

; Patent No. 5830760

; GENERAL INFORMATION:

; APPLICANT: Tsai, S. and S.J. Collins

; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Res

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness

; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue

CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage
COMPUTER: IBM PC/386 Compatible
OPERATING SYSTEM: MS-DOS 4.01
SOFTWARE: Word for Windows 5.01-t
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,383
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/099,242
FILING DATE: July 28, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Broderick, Thomas F.
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: PHCR-1-7190
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2940 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: see Figure 16B; RAR-alpha
IMMEDIATE SOURCE:
LIBRARY: cDNA
US-08-592-383-1

Query Match 85.2%; Score 1777.2; DB 2; Length 2940;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1785; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY	288	CTCCATTGAGACCCAGACGACGTTCTGAAAGAGATAGTGCACGACCTCCCTCCGCCACC	347
DB	279	CACCATTTGAGACCCAGACGACGTTCTGAAAGAGATAGTGCACGACCTCCCTCCGCCACC	338
QY	348	CCCTCTACCCCGCATACAAAGCCTTGCTTGTCTGTCAGGACAGTCTCTCAGGCTACCA	407
DB	339	CCCTCTACCCCGCATACAAAGCCTTGCTTGTCTGTCAGGACAGTCTCTCAGGCTACCA	398
QY	408	CTATGGGTGACGCCCTGTGAGGCTGCAAGGGCTTCTTCGCGCGCAGCATCCAGAAGAA	467
DB	399	CTATGGGTGACGCCCTGTGAGGCTGCAAGGGCTTCTTCGCGCGCAGCATCCAGAAGAA	458
QY	468	CATGGGTACACGTGTACCCGGGACAGAACTGCATCATCAACAAAGGTGACCCGGAACCC	527
DB	459	CATGGGTACACGTGTACCCGGGACAGAACTGCATCATCAACAAAGGTGACCCGGAACCC	518
QY	528	CTGCCACTACTGCCGACTGCAGAAAGTCTTGAAGTGGGATGTCCTCAGAGGAGTCTGTGAG	587
DB	519	CTGCCACTACTGCCGACTGCAGAAAGTCTTGAAGTGGGATGTCCTCAGAGGAGTCTGTGAG	578
QY	588	AAACGCCGAAACAAAGAAAGAGAGGTGCCCAAGCCGAGTCTCTGAGAGCTACAC	647
DB	579	AAACGCCGAAACAAAGAAAGAGAGGTGCCCAAGCCGAGTCTCTGAGAGCTACAC	638
QY	648	GCTGACGCCGAGGTGGGGAGTCAATTGAGAGAGGTGGGCAAGCCGACCAAGAACCTT	707
DB	639	GCTGACGCCGAGGTGGGGAGTCAATTGAGAGAGGTGGGCAAGCCGACCAAGAACCTT	698
QY	708	CCCTGCCCTCTGCCAGTGGGCAAAATACACTACGAAACAAAGCTCAGAACAGTGTCTC	767
DB	699	CCCTGCCCTCTGCCAGTGGGCAAAATACACTACGAAACAAAGCTCAGAACAGTGTCTC	758
QY	768	TCCTGGACATTGACCTCTGGGACAAAGTTGCTGAGTCACTCCACCAAGTGCATTAAGAC	827
DB	759	TCCTGGACATTGACCTCTGGGACAAAGTTGCTGAGTCACTCCACCAAGTGCATTAAGAC	818

QY	828	TGTGGAGTTGCGCAAGACGCTGCCGGCTTTCACACCCCTCACCATCGCCGACAGATCAC	887
DB	819	TGTGGAGTTGCGCAAGACGCTGCCGGCTTTCACACCCCTCACCATCGCCGACAGATCAC	878
QY	888	CTCTCTCAAGGCTGCTGCTGCTGACATCTCTGATCTCTGCGATCTGACGCGGTACAGCC	947
DB	879	CTCTCTCAAGGCTGCTGCTGCTGACATCTCTGATCTCTGCGATCTGACGCGGTACAGCC	938
QY	948	CGAGCAGGACACCATGACCTTCTCGGACGGGCTGACCCCTGAACCGGACCCAGATGCACAA	1007
DB	939	CGAGCAGGACACCATGACCTTCTCGGACGGGCTGACCCCTGAACCGGACCCAGATGCACAA	998
QY	1008	CGCTGGCTTCGGCCCTCTCACCGACCTGGTCTTTGGCTTCGCCAACAGATGTGTCCTT	1067
DB	999	CGCTGGCTTCGGCCCTCTCACCGACCTGGTCTTTGGCTTCGCCAACAGATGTGTCCTT	1058
QY	1068	GGAGATGATGATGCGGAGACGGGCTGCTCAGGCGCATCTGCTCATCTGCGGAGACCG	1127
DB	1059	GGAGATGATGATGCGGAGACGGGCTGCTCAGGCGCATCTGCTCATCTGCGGAGACCG	1118
QY	1128	CCAGGACCTTGAGCAGCCCGGCTGGACATGCTGACGAGCGCTGCTGGAGGCGCT	1187
DB	1119	CCAGGACCTTGAGCAGCCCGGCTGGACATGCTGACGAGCGCTGCTGGAGGCGCT	1178
QY	1188	AAAGTCTACGTGCGGAAGCGGAGCCCGCCCGCCACATGTTCCCAAGATGCTTAAT	1247
DB	1179	AAAGTCTACGTGCGGAAGCGGAGCCCGCCCGCCACATGTTCCCAAGATGCTTAAT	1238
QY	1248	GAAGATTAAGTCTGCGGAGCATGACCGCCCAAGGGGCTGACGCGGTGATCAGCTGAA	1307
DB	1239	GAAGATTAAGTCTGCGGAGCATGACCGCCCAAGGGGCTGACGCGGTGATCAGCTGAA	1298
QY	1308	GATGAGATCCCGGCTCCATGCCGCTCTCATCCAGGAAATGTTGAGAACTCAGAGGG	1367
DB	1299	GATGAGATCCCGGCTCCATGCCGCTCTCATCCAGGAAATGTTGAGAACTCAGAGGG	1358
QY	1368	CCTGACACTCTGAGCGGACACCGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1427
DB	1359	CCTGACACTCTGAGCGGACACCGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1418
QY	1428	CCCGCCAGCAGCTTAGCCCCCAGCTCAGCCCCCAGCTCAGAGAGAGAGAGAGAGAGAGAG	1487
DB	1419	CCCGCCAGCAGCTTAGCCCCCAGCTCAGCCCCCAGCTCAGAGAGAGAGAGAGAGAGAGAG	1478
QY	1488	CCACTCCCGGTGACCGCCACACGACATGAGACACAGCCCTCGCCCTCGCCCGGGCTTTT	1547
DB	1479	CCACTCCCGGTGACCGCCACACGACATGAGACACAGCCCTCGCCCTCGCCCGGGCTTTT	1538
QY	1548	CTCTGCCCTTTTACCGACCATGTGACCCCGCACCAGCCCTGCCCTGCCCTGCCCTGCCG	1607
DB	1539	CTCTGCCCTTTTACCGACCATGTGACCCCGCACCAGCCCTGCCCTGCCCTGCCCTGCCG	1598
QY	1608	GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1667
DB	1599	GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1658
QY	1668	AGGCTGGGCGCTCAGTGGAGTGCCTGCCACAGCCCTGCGCTGACGCTGACGCTGACGCTG	1727
DB	1659	AGGCTGGGCGCTCAGTGGAGTGCCTGCCACAGCCCTGCGCTGACGCTGACGCTGACGCTG	1718
QY	1728	GCCAGGAACTGAGTGGGCGCTGCTGCTGGGTCTCAGATGGGTCTCGGGGGCTTCGTG	1787
DB	1719	GCCAGGAACTGAGTGGGCGCTGCTGCTGGGTCTCAGATGGGTCTCGGGGGCTTCGTG	1778
QY	1788	TTTATCAAGACACCCCTCTGCCAGCTCACACATCTTTCATCACAGCAAAAGCCAGGAC	1847
DB	1779	TTTATCAAGACACCCCTCTGCCAGCTCACACATCTTTCATCACAGCAAAAGCCAGGAC	1838
QY	1848	TTGGCTCCCCCATCTCAGAACTCAGAGCATTTGCTCCCGAGCTGGGGAACTCAACCT	1907
DB	1839	TTGGCTCCCCCATCTCAGAACTCAGAGCATTTGCTCCCGAGCTGGGGAACTCAACCT	1898

Db 1899 CCCCCCTGCTGGTGTGACAGAGGGGTGGGACAGGGGGGTCCCCCTGTAC 1958
Qy 1968 ATACCCTGCCATACCAACCCAGGATTAATTCCTGCTGGTTTGTATTAATTT 2027
Db 1959 ATACCCTGCCATACCAACCCAGGATTAATTCCTGCTGGTTTGTATTAATTT 2018
Qy 2028 TTTTGTGTTGATTTTAAATAAGAAATTTTCATTTTAAGCACAAAAA 2085
Db 2019 TTTTGTGTTGATTTTAAATAAGAAATTTTCATTTTAAGCACATTA 2076

RESULT 9

US-08-592-383-3
; Sequence 3, Application US/08592383
; Patent No. 5830760
; GENERAL INFORMATION:
; APPLICANT: Tsai, S. and S.J. Collins
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Recept
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage
; COMPUTER: IBM PC/386 Compatible
; OPERATING SYSTEM: MS-DOS 4.01
; SOFTWARE: Word for Windows 5.01-t
; CURRENT APPLICATION DATA: US/08/592.383
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/099,242
; FILING DATE: July 28, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Broderick, Thomas F.
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: PHCR-1-7190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2658 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: page 4, RAR-alpha403 dominant negative; deleted of 1311-1596 of SEQ.
; IMMEDIATE SOURCE:
; LIBRARY: cDNA
US-08-592-383-3

Query Match 57.5%; Score 1200; DB 2; Length 2658;
Best Local Similarity 83.5%; Pred. No. 1.2e-233;
Matches 1501; Conservative 0; Mismatches 15; Indels 282; Gaps 1;

Qy 288 CTCATTGAGACCCAGACAGCAGTTCGAAAGAGATAGTGCCACGCCCTCCCTCGCCACC 347
Db 279 CACCATTTGAGACCCAGACAGCAGTTCGAAAGAGATAGTGCCACGCCCTCCCTCGCCACC 338
Qy 348 CCCTCTACCCCGCATCTACAGCCCTTGTCTGTCTGTCAGGACAGTCTCAGGCTACCA 407
Db 339 CCCTCTACCCCGCATCTACAGCCCTTGTCTGTCTGTCAGGACAGTCTCAGGCTACCA 398
Qy 408 CTATGGGCTCAGCGCTGTGAGGGCTGCAAGGGCTTCCTTCGCGCCGACGATCCAGAAGAA 467
Db 399 CTATGGGCTCAGCGCTGTGAGGGCTGCAAGGGCTTCCTTCGCGCCGACGATCCAGAAGAA 458

Qy 468 CATGCTGTACACGTGTACCGGGACAAAGTGCATCATCAACAAGTGCACCCGGAACCC 527
Db 459 CATGGTGTACACGTGTACCGGGACAAAGTGCATCATCAACAAGTGCACCCGGAACCG 518
Qy 528 CTGCCAGTACTGCCGACTGCAGAAAGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAG 587
Db 519 CTGCCAGTACTGCCGACTGCAGAAAGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAG 578
Qy 588 ARAAGCCGCAACAAAGAGAGAGAGTGCACCAAGCGGAGTGCCTGTGAGAGCTACAC 647
Db 579 ARAAGCCGCAACAAAGAGAGAGTGCACCAAGCGGAGTGCCTGTGAGAGCTACAC 638
Qy 648 GCTGACGCGGAGGTGGGGAGCTCATTTGAGAAAGTGCACCAAGCGGAGGAAACCTT 707
Db 639 GCTGACGCGGAGGTGGGGAGCTCATTTGAGAAAGTGCACCAAGCGGAGGAAACCTT 698
Qy 708 CCTGCCCCCTGTGCCAGCTGGGCAAAATACACTAGCAACACAGCTCAGAACAACTGTGCTC 767
Db 699 CCTGCCCCCTGTGCCAGCTGGGCAAAATACACTAGCAACACAGCTCAGAACAACTGTGCTC 758
Qy 768 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTCCACCAAGTGCATCATTAAGAC 827
Db 759 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTCCACCAAGTGCATCATTAAGAC 818
Qy 828 TGTGGAGTTTCGCCAAGCAGCTGCCGGCTTCACACCTCACCACCTCCCGACAGATCAC 887
Db 819 TGTGGAGTTTCGCCAAGCAGCTGCCGGCTTCACACCTCACCACCTCCCGACAGATCAC 878
Qy 888 CCTCTCAAGGCTGCTGCTGGACATCTCTGATCTCTGCGGATCTGCACGGGTTACAGCC 947
Db 879 CCTCTCAAGGCTGCTGCTGGACATCTCTGATCTCTGCGGATCTGCACGGGTTACAGCC 938
Qy 948 CGAGCAGACACCATGACCTTCTCGGACGGGCTGACCTGAAACCGGACCCAGATGCACAA 1007
Db 939 CGAGCAGACACCATGACCTTCTCGGACGGGCTGACCTGAAACCGGACCCAGATGCACAA 998
Qy 1008 CGCTGGCTTCGGCCCTCACCAGCTGGTCTTTCCTTCGCAACACAGCTGCTGCCCT 1067
Db 999 CGCTGGCTTCGGCCCTCACCAGCTGGTCTTTCCTTCGCAACACAGCTGCTGCCCT 1058
Qy 1068 GGAGATGGATGATGCGGAGACGGGCTGCTCAGCGCATCTGCTCATCTCGGAGACCG 1127
Db 1059 GGAGATGGATGATGCGGAGACGGGCTGCTCAGCGCATCTGCTCATCTCGGAGACCG 1118
Qy 1128 CCAGGACCTGGAGCAGCCGAGACCGGTTGGACATGCTGTCAGGAGCCGCTGCTGGAGGCGCT 1187
Db 1119 CCAGGACCTGGAGCAGCCGAGACCGGTTGGACATGCTGTCAGGAGCCGCTGCTGGAGGCGCT 1178
Qy 1188 AAAGTCTACGTCCGGAAGCGGAGGCCAGCCGCCCTCCACATGTTCCCAAGATGCTAAT 1247
Db 1179 AAAGTCTACGTCCGGAAGCGGAGGCCAGCCGCCCTCCACATGTTCCCAAGATGCTAAT 1238
Qy 1248 GAAGATTACTGACCTCGGAAGCATCAGCGCAAGGGGCTGAGCGGCTGATCAGCGCTGAA 1307
Db 1239 GAAGATTACTGACCTCGGAAGCATCAGCGCAAGGGGCTGAGCGGCTGATCAGCGCTGAA 1298
Qy 1308 GATGAGATCCCGGGCTCCATGCCCTCTCATCCAGGAAATGTTGAGAACTCAGAGGG 1367
Db 1299 GATGAGATCCGTTAG----- 1312
Qy 1368 CCTGGACACTGTGAGCGGACAGCCGGGGGTGGGGGGGAGCGGGGTGGCCTGCCCCC 1427
Db 1313 ----- 1312
Qy 1428 CCGCGCAGCAGCTGTAGCCCCCAGCCTCAGCCCCAGCTCCACAGAGACGCGCGCCAC 1487
Db 1313 ----- 1312
Qy 1488 CCACTCCCGTGACCGCCCAAGCCACATGACACAGCCCTCGCCCTCCGCCCGCGCTTTT 1547
Db 1313 ----- 1312
Qy 1548 CTCTGCCTTTTACCGAACCATGTGACCCCGCACAGCCCTGCCCCCACCTGCCCCCTCCCG 1607

QY 945 GCCGAGCAGGACACCATGACCTTCTCGAGGGGTGACCTGAACCGGACCCAGATGCA 1004
DB 1134 CCCAGAACAGACCATGACTTCTCAGACGGCTTACCTTAATCGAATCAGATGCA 1193
QY 1005 CAACGCTGGCTCGGCCCTCCAGGACCTGGTCTTTGCCCTTGGCCAAACAGCTGCTGCC 1064
DB 1194 CAATGCTGGATTGGTCTGACTGACCTTGTGTACCTTTGCCAACACGCTCCTGCC 1253
QY 1065 CCTGAGATGATGATGCGGAGAGGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGA 1124
DB 1254 TTTGGAATGATGACACAGAACAGGCTTCTCAGTCCCATCTGCTTAATCTGTGGAGA 1313
QY 1125 CCGCAGACCTCGGAGCAGCGCCGCGGTGGACATGCTGAGGAGCGCTGCTGGAGGC 1184
DB 1314 CCGCAGGACCTTGAAGAACCGACAAAGTAGATAAGCTACAAGAACCATTTGCTGGAGC 1313
QY 1185 GCTAAAGTCTACCTGGGAGCGGAGCGCCAGCGGCCGCCACATGTTTCCCAAGATGCT 1244
DB 1374 ACTAAAATTTATATCAGAAAAGACGACCCAGCAAGCTCACATGTTTCCAAAGATCTT 1433
QY 1245 ANTGAAGATTACTGACCTGGGAGCATCAGCGCCAGGGGCTCAGCGGTGATCACGCT 1304
DB 1434 AATGAATACAGATCTCGTAGCATGCTTAAAGGTGAGAGCGTGTAAATACCTT 1493
QY 1305 GAAGATGAGATCCCGGCTCCATGCGGCTCTCATCCAGGAAATGTTGGAGAACTCAGA 1364
DB 1494 GAAATGAAATCTCTGGATCAATGCCACCTCTCATTCAGAAATGATGAGAAATCTGA 1553
QY 1365 GGCCTGGACATCTGAGC 1383
DB 1554 AGGACATGAACCTTGACC 1572

RESULT 13

US-08-896-365-5
; Sequence 5, Application US/08896365
; Patent No. 5939264
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Max F.
; APPLICANT: Tuggle, Christopher K.
; APPLICANT: Messer, Lori A.
; APPLICANT: Tun-Ping, Yu
; TITLE OF INVENTION: GENES AND GENETIC MARKERS FOR IMPROVED
; TITLE OF INVENTION: REPRODUCTIVE TRAITS IN ANIMALS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zarley, McKee, Thomte, Voorhees & Sease
; STREET: 801 Grand Avenue, Suite 3200
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/896,365
; FILING DATE: 18-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/022,180
; FILING DATE: 19-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nebel, Heidi S.
; REGISTRATION NUMBER: 37,719
; REFERENCE/DOCKET NUMBER: ISURF 021591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 515-288-3667
; TELEFAX: 515-288-1338
; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 558 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Porcine embryo RARG
; DEVELOPMENTAL STAGE: Embryo
; TISSUE TYPE: Embryo
US-08-896-365-5

Query Match 14.1%; Score 293.4; DB 2; Length 558;
Best Local Similarity 77.2%; Pred. No. 1e-50;
Matches 429; Conservative 0; Mismatches 121; Indels 6; Gaps 6;
QY 649 CTGACGCGGAGGTGGGGAGCTCATTGAGAGGTGGCGAAGCGCCAGGAAAC-CTT 707
DB 3 CTGAGCCCCAGTTAGAGAGCTCATCACCAGGTGACGAAAGTCCATCAAGAGAGGCTC 62
QY 708 CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAAACAGCTCAGAACAAACGCTCTC 767
DB 63 CTCGCCGCTCTGCCAGCTGGGCAATACACCAGCACTCCAGTTAGACCACCGTGTGCA 122
QY 768 TCTGGACATTCACCTCTCGGGA-CAAGTTTCACTGAACTCTCCACCAGTGCATCATTAAGA 826
DB 123 GCTGGATCTGGGGCTGTGGGACCAAGTTCACTGAGCTGGTCAACCAAGTGCATCATTAAGA 182
QY 827 CTGTGGAGTTGCGCAAGCAGCTGCCGGCTTACCACCT-CACCATCGCCGACGATC 885
DB 183 TCGCGGAGTTGTCAAGCGGCTGCTCTTTTACAGGCTCCAGTATTTGCTGACGATC 242
QY 886 ACCCTCTCAAGGCTGCTGCTGGACATCTGTGCTGCGGATCTGCACGCGGTACACG 945
DB 243 ACTCTGCTCAAGGCTGCTGCTGGACATCTGTGCTGCGGNTCTGCACAAAGTTCAC 302
QY 946 CCC-GAGCAGACACCATGACCTTCTCGGAGGGCTGACCTGAACCGGACCCAGATGCA 1004
DB 303 CCGGCGCAGNCTATCATGCACTCTCTGTATGGGTGACCTGAAACCGGACCATGCA 362
QY 1005 CAACGCTGGCTTC-GGCCCCCTCACCAGCTGGTCTTTGCTTCCGCAACGAGCTGCTGC 1063
DB 363 CGATGCTGACTTCGGGCCCCCTCACAGACNTCTGTGTGCTTGTGCTGGCAGCTCCTGC 422
QY 1064 CCCTGGAGATGGATGATGCGGAGACGGGCTGCTCAG-CGCCATCTGCTCATCTGCGGA 1122
DB 423 CACTGGAGATGGATGACAGAGACAGGCTGCTCAGCGCCATCTGCTCATCTGCTGCGGA 482
QY 1123 GACCGCAGGACCTGGAGCAGCGGCTGCTGACATGCTGACAGGACCGCTGCTGGAG 1182
DB 483 GACCGCATGACCTGGAGGAAACCGAGTAAAGTGACAGCTGCAGGAGCCATTGCTGGA 542
QY 1183 GCCTAAAGTCTACG 1198
DB 543 GCCTGAGGCTCTATG 558

RESULT 14

US-08-592-383-6
; Sequence 6, Application US/08592383
; Patent No. 5830760
; GENERAL INFORMATION:
; APPLICANT: Tsai, S. and S.J. Collins
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Re
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA

ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage
COMPUTER: IBM PC/386 Compatible
OPERATING SYSTEM: MS-DOS 4.01
SOFTWARE: Word for Windows 5.01-t
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,383
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/099,242
FILING DATE: July 28, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Broderick, Thomas F.
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: FHC-1-7190
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: page 11, RAR-alpha DNA binding region; positions 364 to 564 of SEQ.
IMMEDIATE SOURCE:
LIBRARY: CDNA
US-08-592-383-6

Query Match 9.6%; Score 199.4; DB 2; Length 201;
Best Local Similarity 99.5%; Pred. No. 7.5e-32;
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 373 TGTCTTGTCTGTAGGACAGTCTCTAGGCTACCACTATGGGGTCAGCGCCCTGTGAGGGC 432
Db 1 TGTCTTGTCTGTAGGACAGTCTCTAGGCTACCACTATGGGGTCAGCGCCCTGTGAGGGC 60
QY 433 TGAAGGGCTTCTCCCGCCGAGCATCCAGAAACATGGTGTACACGTGTACCGGGAC 492
Db 61 TGNAGGGCTTCTCCCGCCGAGCATCCAGAAACATGGTGTACACGTGTACCGGGAC 120
QY 493 AAGAACTGCATCATCAACAAGGTGACCGGAAACCCCTGCCAGTACTGCCGACTGCAGAA 552
Db 121 AAGAACTGCATCATCAACAAGGTGACCGGAAACCCCTGCCAGTACTGCCGACTGCAGAA 180
QY 553 TGCTTTGAAGTGGCGATGTCC 573
Db 181 TGCTTTGAAGTGGCGATGTCC 201

RESULT 15
US-08-336-408B-1
Sequence 1, Application US/08336408B
Patent No. 5723329
GENERAL INFORMATION:
APPLICANT: EVANS, RONALD M.
APPLICANT: MANGELSDORF, DAVID J.
TITLE OF INVENTION: RETINOID RECEPTOR COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
STREET: 444 SOUTH FLOWER STREET, SUITE 2000
CITY: LOS ANGELES
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/336,408B
APPLICATION NUMBER: US/08/336,408B
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/933,453
FILING DATE: 21-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US91/00399
FILING DATE: 22-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/478,071
FILING DATE: 09-FEB-1990
ATTORNEY/AGENT INFORMATION:
NAME: REITER, STEPHEN E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9851
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-1995
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1866 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 76...1464
US-08-336-408B-1

Query Match 8.8%; Score 183; DB 1; Length 1866;
Best Local Similarity 50.6%; Pred. No. 3e-28;
Matches 537; Conservative 0; Mismatches 500; Indels 24; Gaps 3;
QY 239 CCCAGAGCAGCAGTCTCTGAAGAGATAGTCCCGCAGCCCTCCCTCGCCACCCCTCTACCCC 358
Db 404 CCTGGGGCTCAATGGCGCTCTCAAGGTCCCGCCGCCACCCCTCAGGAACAGTGGCTTCT 463
QY 359 GCATCTACAAGCTTGTCTGTCTGTCAGGACAAAGTCTCTCAGCTACCACTATGGGTCA 418
Db 464 TCACCAAGCACATCTGCGCCATCTGCGGGGACCGCTCTCAGGCAAGCAGTGGAGTGT 523
QY 419 GCGCTGTGAGGGCTGCAAGGGCTTCTCCCGCGGAGCATCCAGAAAGACATGGTGTACA 478
Db 524 ACAGCTGCGAGGGTGCAGGGCTTCTTCAAGCGGACGGTGCAGCAAGGACCTGACCTACA 583
QY 479 CGTGTACCGGGGACAAAGAACTGCATCATCAACAAGGTGACCGGAAACCCCTGCCAGTACT 538
Db 584 CCTGCCGCGACAAACAGGACTGCTGATTCAGCAAGCGGACCGGAAACCCCTGCCAGTACT 643
QY 539 GCCGACTGCAGAGTGTCTTGAAGTGGGATGTCTCAAGAGAGTCTGTGAGAAACAGCCGAA 598
Db 644 GCGCTACCAAGAGTGTCTGCGCATGGGATGAAGCGGGAAGCCGTGCAGGAGAGCG -- 701
QY 599 ACAAGAAGAAAGAGGTGCCCAAGCCCGAGTGTCTTGAGAGCTACACGCTGACGCCGG 658
Db 702 -GACGCTGGCAAGGACCGGAAAGAGTGGAGTGGAGTGCAGCAGCGGCAAGCAGG 760
QY 659 AGGTGGGGAGTCTATGAGAAGGTGCGCAAGCGGACAGGAAACCTTCCCTGCCCTCT 718
Db 761 ACATGCGGTGGAGAGGATCTCTGAGGCTGAGCTGCGGCTGGAGCGCAGACGAGACCT 820
QY 719 GCCAGTGGGCAATACACTAGCAACAACAGCTCAGAAACACGCTCTCTCTGACATTTG 778
Db 821 ACCTGTGGCAAGTTCAGTGAATCTTCCACCAAGTTCATTAAGACTGTGGAGTTCG 838
QY 863 ACCCTGTACCAACATTTGCCAAGCAGCGCAGCAACAGCTTTTCCACCTGTGGAGTGGG 922

QY 839 CCAAGCAGCTGCCCGGCTTACACACCTCACCATCGCGGACCATGATCACCCTCTCAAGG 898
 Db 923 CCAAGCGGATCCCACTTCTCAGAGTGCCTCGACACCAAGGTCTATGCTGCGGG 982
 QY 899 CTGCTGCTGGGATCTGATCTGCGGATCTGACGCGGTACACGCCGAGCAGACA 958
 Db 983 CAGCTGGAATGAGTGTCTATGCCCTCTTCCCAACCGCTCATCGCGTGAAGGACG 1042
 QY 959 CCATGACCTTCTCGGAGCGGCTGACCTGAACCGGACCCAGATGCACAAGGCTGGCTCG 1018
 Db 1043 GGATCTCTTGGCCACCGGCTGACGCTCCACCGGACACGCCGCCAGCGGAGTGG 1102
 QY 1019 GCCCTCTACCGACCTGGTCTT---TGCCTTCGCCAACACAGCTGCTGCCCTGGAGATGG 1075
 Db 1103 GCCTCATCTTTGACAGGGTCTGACGAGCTTGTGCCAAGATCGGGACATGCAGATGG 1162
 QY 1076 ATGATCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTCGGGAGACGCCAGGACC 1135
 Db 1163 ACAGAGGAGCTGGGCTGCTCGCGCCATCTGCTCTTTAAACCTGACTCCAAAGGGC 1222
 QY 1136 TGGAGACCGCGACCGGCTGGACATGCTCAGGAGCGCTGCTGAGGCGCTAAAGGTCT 1195
 Db 1223 TCTCGAACCCGCGAGGTGGAGCGCTGAGGAGAGGTCTATGCTGCTTGGAGCCT 1282
 QY 1196 AGTGGGAGAGCGGAGCGCCAGCGCCGCCACATGTTCCCAAGATGCTAATGAAGATTA 1255
 Db 1283 ACTCAAGCACAAAGTACCCAGAGCAGCGGGAAGTTGCTAAGCTCTTCTCGGCTGC 1342
 QY 1256 CTGACCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGATCAGCTGAAGATGGAGA 1315
 Db 1343 CGGCTTGGGCTCCATCGGGGCTCAAAATGCTGGAACATCTCTTCTTCAAGCTCATCG 1402
 QY 1316 TCCCGGGCTCCATGCGGCTCTCATCCAGAAATGTTGGAG 1356
 Db 1403 GGGACACACCATTGACACCTTCTTATGGAGATGCTGGAG 1443

Search completed: March 29, 2003, 15:19:14
 Job time : 104 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 09:44:04 ; Search time 390 seconds
(without alignments)
12045.304 Million cell updates/sec

Title: US-09-691-220-1
Perfect score: 2086
Sequence: 1 aacagcagcgagggcgagggg.....cacaaaaa.....2086

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2086	100.0	2086	24	cdna encoding huma
2	1782.2	85.4	3036	13	myl/RAR-alpha fusi
3	1782.2	85.4	3036	19	Human PML/RARalpha
4	1782.2	85.4	3036	20	Human myl/RAR-alpha
5	1782.2	85.4	3511	21	Human PML/RAR-alpha
6	1779	85.3	3511	17	PML-RAR-alpha DNA
7	1778.8	85.3	2907	24	Human cdna differe
8	1778.8	85.3	2907	24	Lung cancer relate
9	1778.8	85.3	2928	20	Human RAR-alpha cd

10	1777.4	85.2	3511	17	AAU33246	PML-RAR-alpha RNA
11	1777.2	85.2	2928	13	AAQ29338	RAR-alpha gene. S
12	1777.2	85.2	2940	10	AAH90124	DNA of clone phRAR
13	1777.2	85.2	2940	16	AAQ81476	RAR-alpha clone la
14	1707.4	81.9	3052	23	AA583047	DNA encoding novel
15	1540.2	73.8	2930	23	AA583048	DNA encoding novel
16	1500.8	71.9	2073	23	AA585135	DNA encoding novel
17	1500.8	71.9	2457	23	AA583049	DNA encoding novel
18	1499.4	71.9	1944	23	AA583045	DNA encoding novel
19	1499.4	71.9	1944	23	AA585136	DNA encoding novel
20	1207.6	57.9	1692	23	AA583808	DNA encoding novel
21	1200	57.5	2658	16	AAQ81477	RAR-alpha-403 domi
22	1089.2	52.2	1481	17	AAU33247	RAR-alpha RNA sequ
23	1086	52.1	1481	17	AAU33260	RAR-alpha RNA sequ
24	749.2	35.9	1576	12	AAQ10226	Clone pGEM-hRARgam
25	747.6	35.8	1903	12	AAQ10389	Encodes human Reti
26	747.6	35.8	1903	16	AAU05196	Human retinoic aci
27	715.2	34.3	2521	13	AAQ20542	Mouse retinoic aci
28	715.2	34.3	2740	13	AAQ20541	Mouse retinoic aci
29	713.6	34.2	1913	16	AAU05202	Murine retinoic ac
30	710.4	34.1	1912	12	AAQ10388	Encodes skin-speci
31	701.2	33.6	1375	24	ABK37493	Human FLAG RARbeta
32	701.2	33.6	1399	24	ABK37492	Human FLAG RARbeta
33	701.2	33.6	8887	24	ABK37494	Vector genome plas
34	701.2	33.6	9151	24	ABK37495	Vector genome plas
35	674	32.3	2992	16	AAQ79934	Human liver hap cd
36	674	32.3	2992	17	AAU06491	Human liver hap cd
37	672.4	32.2	2992	15	AAQ65572	Human hepatoma ret
38	667.8	32.0	2989	10	AAH90093	Human liver hap cd
39	648.8	31.1	20512	24	AAU38339	Human Hap (hepatom
40	591.4	28.4	595	24	AA562827	Genomic DNA encodi
41	552	26.5	1351	16	AAQ79935	cdna sequence #614
42	296.2	14.2	5838	22	AAH98429	Human liver RAR-be
43	272.8	13.1	301	22	AAU79992	Human EST-derived
44	272.6	13.1	482	23	AA581232	Nucleotide sequenc
45	271.6	13.0	1641	23	AA586582	DNA encoding novel

ALIGNMENTS

RESULT 1
AAL38338
ID AAL38338 standard; cdna; 2086 BP.
AC AAL38338;
XX
XX
DT 15-AUG-2002 (first entry)
XX
DE cdna encoding human nuclear hormone receptor protein.
XX
DE Human nuclear hormone receptor; gene therapy;
KW immune response; gene; ss.
XX
XX Homo sapiens.
OS.
FH Key Location/Qualifiers
CDS 127..1500
FT /*tag=
FT /product= "Human nuclear hormone receptor protein"
XX
XX WO200231146-A2.
XX
XX PD 18-APR-2002.
XX
XX PF 05-OCT-2001; 2001WO-US31095.
XX
XX PR 11-OCT-2000; 2000US-239117P.
XX
XX PR 19-OCT-2000; 2000US-0691220.
XX
XX PA (PEKE) PE CORP NY.
XX
XX PI Wei M, Ye J, Yan C, Ketchum KA, Di Francesco V, Beasley EM;

XX WPI: 2002-426282/45.
DR P-PSDB; AAO21489.
XX
PT New human nuclear hormone receptor proteins and nucleic acids, useful
PT as models or targets for developing human therapeutic targets, and in
PT identifying therapeutic proteins and modulators of nuclear hormone
PT receptor expression
XX
PS Claim 1; Fig 1; 73pp; English.
XX
CC The invention relates to an isolated peptide of a novel human nuclear
CC hormone receptor with a fully defined sequence of 457 amino acids given
CC in the specification. The novel human nuclear hormone receptor peptides
CC and nucleic acids encoding them can be used as models for the development
CC of human therapeutic targets, aid in the identification of therapeutic
CC proteins, and serve as targets for the development of human therapeutic
CC agents that modulate nuclear hormone receptor activity in cells and
CC tissues that express the nuclear hormone receptor. The nucleic acids may
CC be used as a query sequence to perform searches against sequence
CC databases to identify family members or related sequences, as probes or
CC primers, to construct recombinant vectors, to identify compounds that
CC modulate nuclear hormone receptor nucleic acid expression, in gene
CC therapy, and as antisense constructs to control nuclear hormone receptor
CC gene expression in cells, tissues or organisms. The polypeptides can be
CC used to raise antibodies or to elicit an immune response, as a reagent in
CC assays designed to determine protein levels in biological fluids, as
CC markers for tissues in which a corresponding protein is expressed, to
CC identify a binding partner/ligand to develop a system for the
CC identification of inhibitors of the binding reaction, in drug screening
CC assays, and to identify compounds that modulate protein activity. This
CC polynucleotide sequence represents the cDNA encoding the human nuclear
CC hormone receptor protein of the invention.
XX
SQ Sequence 2086 BP; 437 A; 685 C; 583 G; 381 T; 0 other;

Query Match 100.0%; Score 2086; DB 24; Length 2086;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACAGCAGAGGCGGAGGAGCTCTCTCCGCCAGCTGCTGTGCGATGGGCGG 60
DB 1 AACAGCAGAGGCGGAGGAGCTCTCTCTCCGCCAGCTGCTGTGCGATGGGCGG 60
QY 61 CCGGCTGAGTGAGCGGGGCGGCGAGGACTTCCAGCTCGGACCTTTCCTTCGAGG 120
DB 61 CCGGCTGAGTGAGCGGGGCGGCGAGGACTTCCAGCTCGGACCTTTCCTTCGAGG 120
QY 121 GGAAGATGTACGAGAGTGTAAGTGGGGGTCCACCCCTAATCCCTTCCTAGTGGTG 180
DB 121 GGAAGATGTACGAGAGTGTAAGTGGGGGTCCACCCCTAATCCCTTCCTAGTGGTG 180
QY 181 GATTTTATTAACAGAACCGGGCTGTTTCTCCAGAGAGGGGCTCCCGCCCGGGT 240
DB 181 GATTTTATTAACAGAACCGGGCTGTTTCTCCAGAGAGGGGCTCCCGCCCGGGT 240
QY 241 CCGTACTCCACCGCTCCGAGTCCGCTTTGGAATGGCTCAACCACTCCATAGAC 300
DB 241 CCGTACTCCACCGCTCCGAGTCCGCTTTGGAATGGCTCAACCACTCCATAGAC 300
QY 301 CAGAGCAGCAGTTCTGAAGATAGTGCCAGCCCTCCCTCGCACCCCTCTACCCCGC 360
DB 301 CAGAGCAGCAGTTCTGAAGATAGTGCCAGCCCTCCCTCGCACCCCTCTACCCCGC 360
QY 361 ATCTACAAGCCTTGTCTGTCTAGGACAAGTCTCAGGCTACCACCTATGGGGTCAGC 420
DB 361 ATCTACAAGCCTTGTCTGTCTAGGACAAGTCTCAGGCTACCACCTATGGGGTCAGC 420
QY 421 GCTGTGAGGGCTGCAAGGGCTTCTTCGCGCAGCATCCAGAAGAACATGGTGTACAG 480
DB 421 GCTGTGAGGGCTGCAAGGGCTTCTTCGCGCAGCATCCAGAAGAACATGGTGTACAG 480
QY 481 TGTACCGGGACAAAGTGCATCATCAACAGGTGACCGGGAACCCCTGCCAGTACTGC 540

DB 481 TGTACCGGGACAAAGTGCATCATCAACAGGTGACCCGGAACCCCTGCCAGTACTGC 540
QY 541 CGACTGCAGAAAGTGTCTTGAAGTGGGCATGTCCAAAGGAGTCTGTGAGAAACACCCGAAAC 600
DB 541 CGACTGCAGAAAGTGTCTTGAAGTGGGCATGTCCAAAGGAGTCTGTGAGAAACACCCGAAAC 600
QY 601 AAGAAGAAGAGGAGGTGCCCAAGCCGAGTCTCTGAGAGCTACACGCTGACGCCGGAG 660
DB 601 AAGAAGAAGAGGAGGTGCCCAAGCCGAGTCTCTGAGAGCTACACGCTGACGCCGGAG 660
QY 661 GTGGGGAGCTCATTTGAAGTGGCAAGCGCACACGAGAAACCTTCCCTGCCCTCTGC 720
DB 661 GTGGGGAGCTCATTTGAAGTGGCAAGCGCACACGAGAAACCTTCCCTGCCCTCTGC 720
QY 721 CAGCTGGGCAATACACTACGAAACACAGCTCAGAACACAGTGTCTCTGAGACATTGAC 780
DB 721 CAGCTGGGCAATACACTACGAAACACAGCTCAGAACACAGTGTCTCTGAGACATTGAC 780
QY 781 CTCTGGGCAAGTTCAAGTGAACCTCTCCACCAAGTGCATCATTAAGACTGTGAGTTCGC 840
DB 781 CTCTGGGCAAGTTCAAGTGAACCTCTCCACCAAGTGCATCATTAAGACTGTGAGTTCGC 840
QY 841 AAGCAGCTGCGCGGCTTCCACCAACCTCACCATCGCGCACAGATCACCTCTCAAGGCT 900
DB 841 AAGCAGCTGCGCGGCTTCCACCAACCTCACCATCGCGCACAGATCACCTCTCAAGGCT 900
QY 901 GCCTGCTGGACATCTCTGCGGATCTGCACGCGGTACACGCCGAGCAGGACAC 960
DB 901 GCCTGCTGGACATCTCTGCGGATCTGCACGCGGTACACGCCGAGCAGGACAC 960
QY 961 ATGACCTTCTCGGAGGGCTGACCTGAAACGAGCCAGATGACACAGCTGGCTTCGC 1020
DB 961 ATGACCTTCTCGGAGGGCTGACCTGAAACGAGCCAGATGACACAGCTGGCTTCGC 1020
QY 1021 CCCCTACCGAGCTGCTCTTTGCTTCCGCAACAGCTGCTGCCCTCGGAGATGATGAT 1080
DB 1021 CCCCTACCGAGCTGCTCTTTGCTTCCGCAACAGCTGCTGCCCTCGGAGATGATGAT 1080
QY 1081 GCGAGAGCGGGGCTGCTCAGCGCCATCTGCTCATCTCGGAGACCGCAGACCTGGAG 1140
DB 1081 GCGAGAGCGGGGCTGCTCAGCGCCATCTGCTCATCTCGGAGACCGCAGACCTGGAG 1140
QY 1141 CAGCCGAGCGGGTGGACATCTGCAGAGCGGCTGCTGGAGGCGCTAAAGCTCTACGTG 1200
DB 1141 CAGCCGAGCGGGTGGACATCTGCAGAGCGGCTGCTGGAGGCGCTAAAGCTCTACGTG 1200
QY 1201 CGGAAGCGGAGGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGAAGATTACTGAC 1260
DB 1201 CGGAAGCGGAGGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGAAGATTACTGAC 1260
QY 1261 CTGCGAAGCATCAGCGCCAGGGGGCTGAGCGGGTGTACGCTGAAGATGGAGATCCG 1320
DB 1261 CTGCGAAGCATCAGCGCCAGGGGGCTGAGCGGGTGTACGCTGAAGATGGAGATCCG 1320
QY 1321 GGCTCCATCGGCTCTCATCCAGGAATGTTGGAGAACTCAGAGGGCTGGACACTCTG 1380
DB 1321 GGCTCCATCGGCTCTCATCCAGGAATGTTGGAGAACTCAGAGGGCTGGACACTCTG 1380
QY 1381 AGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGGCTTGCCTGCCCGCCGAGGACG 1440
DB 1381 AGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGGCTTGCCTGCCCGCCGAGGACG 1440
QY 1441 TGTAGCCCCAGCCTCAGCCCCAGCTCCAACAGAAAGACGCCCGGACACCTACCCCGTGA 1500
DB 1441 TGTAGCCCCAGCCTCAGCCCCAGCTCCAACAGAAAGACGCCCGGACACCTACCCCGTGA 1500
QY 1501 CCGGCCACGCCCATGGACACAGCCCTCGGCCCTCCGCCCGGGCTTTTCTCTGCTCTTCTA 1560
DB 1501 CCGGCCACGCCCATGGACACAGCCCTCGGCCCTCCGCCCGGGCTTTTCTCTGCTCTTCTA 1560
QY 1561 CCGNACATGTGACCCCGCAGCAGCCCTGCCCGCCAGCTGCCCTCCCGGGCAGTACTGGGA 1620

Db 1910 AGCAGGACACCATGACCTTCGCGAGGGCTGACCTGAACCGGACCCAGATGCACAAGC 1969
QY 1010 CTGGCTTCGGGCCCCCTACCGACCTGGTCTTTGCTTCGCCCAACACAGCTGTCGCCCTGG 1069
Db 1970 CTGGCTTCGGGCCCCCTACCGACCTGGTCTTTGCTTCGCCCAACACAGCTGTCGCCCTGG 2029
QY 1070 AGATGGATGATCGGAGACGGGGTGTCTACGGCCATCTGCTCATCTCGGAGACCGCC 1129
Db 2030 AGATGGATGATCGGAGACGGGGTGTCTACGGCCATCTGCTCATCTCGGAGACCGCC 2089
QY 1130 AGGACCTGGAGCAGCGGACCGGTGGACATGCTGCAGGAGCGCTGCTGGAGCGCTAA 1189
Db 2090 AGGACCTGGAGCAGCGGACCGGTGGACATGCTGCAGGAGCGCTGCTGGAGCGCTAA 2149
QY 1190 AGGTCTACGTGCGGAAGCGGAGCGCCAGCGCCGCCACATGTTCCCAAGATGCTAAGA 1249
Db 2150 AGGTCTACGTGCGGAAGCGGAGCGCCAGCGCCGCCACATGTTCCCAAGATGCTAAGA 2209
QY 1250 AGATTACTGACCTCGGAAGCATAGCGCCCAAGGGGGTGTAGCGGTGATCAGCTGAAGA 1309
Db 2210 AGATTACTGACCTCGGAAGCATAGCGCCCAAGGGGGTGTAGCGGTGATCAGCTGAAGA 2269
QY 1310 TGGAGATCCCGGGTCTCATCGCGCTCTCATCAGGAAATTTGGAGAACTCAGAGGGCC 1369
Db 2270 TGGAGATCCCGGGTCTCATCGCGCTCTCATCAGGAAATTTGGAGAACTCAGAGGGCC 2329
QY 1370 TGGACACTCTGAGGGGACGCGGGGGTGGGGGGGAGCGGGGTGGCTGCCCCCCC 1429
Db 2330 TGGACACTCTGAGGGGACGCGGGGGTGGGGGGGAGCGGGGTGGCTGCCCCCCC 2389
QY 1430 CGCCAGGACGTGTAGCCCGACGCTCAGCCCGCAGCTCCACAGAGCAGCGCCGCCACCC 1489
Db 2390 CGCCAGGACGTGTAGCCCGACGCTCAGCCCGCAGCTCCACAGAGCAGCGCCGCCACCC 2449
QY 1490 ACTCCCGGTACCGCCCGACGCTCAGCCCGCAGCTCCACAGAGCAGCGCCGCCACCC 1549
Db 2450 ACTCCCGGTACCGCCCGACGCTCAGCCCGCAGCTCCACAGAGCAGCGCCGCCACCC 2509
QY 1550 CTGCTTTCTACCGACCATGTGACCCCGCAGCGCTGCCCCACCTGCCCCCGGGC 1609
Db 2510 CTGCTTTCTACCGACCATGTGACCCCGCAGCGCTGCCCCACCTGCCCCCGGGC 2569
QY 1610 AGTACTGGGACCTTCCCTGGGGAGCGGGGAGGAGGAGCGAGCTCCTTGACACAG 1669
Db 2570 AGTACTGGGACCTTCCCTGGGGAGCGGGGAGGAGGAGGAGCGAGCTCCTTGACACAG 2629
QY 1670 GCCTGGGCGCTCAGTGGACTGCTGCTCCACAGCCTGGGCTCAGAGGCGCGAGGC 1729
Db 2630 GCCTGGGCGCTCAGTGGACTGCTGCTCCACAGCCTGGGCTCAGAGGCGCGAGGC 2689
QY 1730 CAGGAACCTGAGTGAGCGCCCTGCTCTGGGTCTCAGATGGGTCTGGGGGCTCGGT 1789
Db 2690 CAGGAACCTGAGTGAGCGCCCTGCTCTGGGTCTCAGATGGGTCTGGGGGCTCGGT 2749
QY 1790 CATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAAAAGCGCAGACT 1849
Db 2750 CATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAAAAGCGCAGACT 2809
QY 1850 GGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 1909
Db 2810 GGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 2869
QY 1910 CCCCTGCTCGGTGGTGACAGAGGGGGTGGACAGGGGGGGGGTCCCCCTGTACAT 1969
Db 2870 CCCCTGCTCGGTGGTGACAGAGGGGGTGGACAGGGGGGGGGTCCCCCTGTACAT 2929
QY 1970 ACCCTGCCATACCAACCCAGGTATTAATCTCGCTGGTGTGTTTTTAATTTTT 2029
Db 2930 ACCCTGCCATACCAACCCAGGTATTAATCTCGCTGGTGTGTTTTTAATTTTT 2989
QY 2030 TTGTTTTGATTTTTTAAAGAAATTTTTCATTTTAAAGCAAAAAA 2076
Db 2990 TTGTTTTGATTTTTTAAAGAAATTTTTCATTTTAAAGCAAAAAA 3036

RESULT 3

AAV20474

ID AAV20474 standard; DNA; 3036 BP.

XX AC AAV20474;

XX DT 17-JUN-1998 (first entry)

XX DE Human PML/RARalpha oncogene.

XX KW Human; oncogene; proto-oncogene; neoplastic disease; anticancer;
XX KW cancer; antisense oligonucleotide; PML/RARalpha; ds.

XX OS Homo sapiens.

XX PN US5734039-A.

XX PD 31-MAR-1998.

XX PF 15-SEP-1994; 94US-0306691.

XX PR 15-SEP-1994; 94US-0306691.

XX PA (UYJE-) UNIV JEFFERSON THOMAS.

XX PI Calabretta B, Skorski T;

XX DR WPI; 1998-229882/20.

XX PT Anticancer composition comprising two anti-sense oligo:nucleotide(s)
XX PS - targeting cytoplasmic and nuclear oncogene(s)

XX CC Claim 1; Column 145-150; 92pp; English.

XX CC The present sequence represents an oncogene from the present invention.
XX CC The present invention describes a composition which comprises two
XX CC antisense oligonucleotides. The first oligonucleotide is specific for a
XX CC cytoplasmic oncogene or proto-oncogene selected from ras, raf, bfgf-1,
XX CC c-fms, c-ros, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-fgr and
XX CC c-yes. The second oligonucleotide is specific for a nuclear oncogene or
XX CC proto-oncogene selected from myc, jun, c-ets, c-fos, c-myc, B-myb,
XX CC c-rel, c-vav, c-ski, c-spl, cyclin D1, PML/RAR alpha, AML1/MTG8,
XX CC E2A/p1 and ALL-1/AF-4. The composition is used for treating cancer.
XX CC The combination of antisense oligonucleotides has synergistically
XX CC enhanced ability to inhibit growth of cancer cells.

XX SQ Sequence 3036 BP; 605 A; 1018 C; 890 G; 523 T; 0 other;

Query Match 85.4%; Score 1782.2; DB 19; Length 3036;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 290 CCATTGAGACCCAGAGCAGCAGTCTTGAAGATAGTGCCCGACCTCCCTCGCCACCCC 349

Db 1250 CCATTGAGACCCAGAGCAGCAGTCTTGAAGATAGTGCCCGACCTCCCTCGCCACCCC 1309

QY 350 CTCTACCCCGCATCTACAAGCCTTGCTTGTCTCAGGACAAGTCTCAGGCTACCACT 409

Db 1310 CTCTACCCCGCATCTACAAGCCTTGCTTGTCTCAGGACAAGTCTCAGGCTACCACT 1369

QY 410 ATGGGTTCAGCGCTGTGAGGGCTTGAAGGGCTTCTTCGCCCGCAGCATCCAGAGAACA 469

Db 1370 ATGGGTTCAGCGCTGTGAGGGCTTGAAGGGCTTCTTCGCCCGCAGCATCCAGAGAACA 1429

QY 470 TGGTGTACAGTGTACCGGGACAAGAACTGCATCATCAAGAGTGCACCGGAACCCCT 529

Db 1430 TGGTGTACAGTGTACCGGGACAAGAACTGCATCATCAAGAGTGCACCGGAACCCCT 1489

QY 530 GCCAGTACTGCCAGCTGCAGAAAGTGTGTTGAAGTGGGCATGTCCAAAGAGTCTGTGAAA 589

Db 1490 GCCAGTACTGCCAGCTGCAGAAAGTGTGTTGAAGTGGGCATGTCCAAAGAGTCTGTGAAA 1549

```
QY 590 ACGACCGAACAAGAGAGAGAGGTGCCCAAGCCCCAGTGGCTCTGAGAGCTACACGC 649
Db 1550 ACGACCGAACAAGAGAGAGAGGTGCCCAAGCCCCAGTGGCTCTGAGAGCTACACGC 1609
QY 650 TGACCCCGAGGTGGGGAGCTCATTGAGAAGTGGCAAGCGCACAGGAAACCTTCC 709
Db 1610 TGACCCCGAGGTGGGGAGCTCATTGAGAAGTGGCAAGCGCACAGGAAACCTTCC 1669
QY 710 CTGCCCTCTGCCAGCTGGGCAAAATACATACGAACAACAGCTCAGAACACCTGTCTCTC 769
Db 1670 CTGCCCTCTGCCAGCTGGGCAAAATACATACGAACAACAGCTCAGAACACCTGTCTCTC 1729
QY 770 TGGACATTGACCTCTGGGCAAGTTGAGTGAAGTCTCCACCAAGTGCATTAAGACTG 829
Db 1730 TGGACATTGACCTCTGGGCAAGTTGAGTGAAGTCTCCACCAAGTGCATTAAGACTG 1789
QY 830 TGGAGTTGCCAAGCAGCTGCCCGGCTTCAACCAACCTCAACATCGCGACCAAGTCAACC 889
Db 1790 TGGAGTTGCCAAGCAGCTGCCCGGCTTCAACCAACCTCAACATCGCGACCAAGTCAACC 1849
QY 890 TCCTCAAGGCTGCCCTGGACATCTCTGATCTCGGATCTGCACGCGGTACACGCCCG 949
Db 1850 TCCTCAAGGCTGCCCTGGACATCTCTGATCTCGGATCTGCACGCGGTACACGCCCG 1909
QY 950 AGCAGGACACCATGACCTTCTGGAGCGGCTGACCCCTGAACCGGACCCAGATGCAACAAG 1009
Db 1910 AGCAGGACACCATGACCTTCTGGAGCGGCTGACCCCTGAACCGGACCCAGATGCAACAAG 1969
QY 1010 CTGGCTCGGCCCCCTACCGACCTGGTCTTTGGCTTGGCCCTTGGCCAAACAGCTGCTGCCCTGG 1069
Db 1970 CTGGCTCGGCCCCCTACCGACCTGGTCTTTGGCTTGGCCCTTGGCCAAACAGCTGCTGCCCTGG 2029
QY 1070 AGATGGATGATCGGAGACGGGGCTGCTCAGCGCCATCTGCTCATCTCTCGGAGACCGCC 1129
Db 2030 AGATGGATGATCGGAGACGGGGCTGCTCAGCGCCATCTGCTCATCTCTCGGAGACCGCC 2089
QY 1130 AGGACCTGGAGACGGCGGCTGGACATGCTCGAGAGCGCGTGTCTGGAGGCGCTAA 1189
Db 2090 AGGACCTGGAGACGGCGGCTGGACATGCTCGAGAGCGCGTGTCTGGAGGCGCTAA 2149
QY 1190 AGGTCTAGTGGGAAGCGGAGGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAATGA 1249
Db 2150 AGGTCTAGTGGGAAGCGGAGGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAATGA 2209
QY 1250 AGATTACTGACCTCGGAAGCATCAGCGCCAGGGGGCTGAGCGGTGTATCACTGCTGAAGA 1309
Db 2210 AGATTACTGACCTCGGAAGCATCAGCGCCAGGGGGCTGAGCGGTGTATCACTGCTGAAGA 2269
QY 1310 TGGAGATCCCGGGTTCATCGCGCTCTCATCCAGGAATGTTGGAGAACTCAGAGGGCC 1369
Db 2270 TGGAGATCCCGGGTTCATCGCGCTCTCATCCAGGAATGTTGGAGAACTCAGAGGGCC 2329
QY 1370 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGGAGCGGGGTGGCCCTGCCCGCCC 1429
Db 2330 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGGAGCGGGGTGGCCCTGCCCGCCC 2389
QY 1430 CGCCAGGAGCTGTAGCCCGCCAGCTCAGCGCCAGCTTCCAAAGAGAGCGCCCGCCACCC 1489
Db 2390 CGCCAGGAGCTGTAGCCCGCCAGCTCAGCGCCAGCTTCCAAAGAGAGCGCCCGCCACCC 2449
QY 1490 ACTCCCGGTGACCGCCACGACATGACAGCCCTCGCCCTCGCGCCCGCGCTTTCT 1549
Db 2450 ACTCCCGGTGACCGCCACGACATGACAGCCCTCGCCCTCGCGCCCGCGCTTTCT 2509
QY 1550 CTGCCCTTCTACGACCATGTGACCCCGCCAGCGCCCTGCGCCCGCCAGCTGCGCCCTCCGGGC 1609
Db 2510 CTGCCCTTCTACGACCATGTGACCCCGCCAGCGCCCTGCGCCCGCCAGCTGCGCCCTCCGGGC 2569
QY 1610 AGTACTGGGACCTTCCCTGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669
Db 2570 AGTACTGGGACCTTCCCTGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2629
```

```
QY 1670 GCCTGGGCGCTCAGTGGACTGCTGCTCCACAGCCCTGGGCTGACGTCAGAGGCGGAGGC 1729
Db 2630 GCCTGGGCGCTCAGTGGACTGCTGCTCCACAGCCCTGGGCTGACGTCAGAGGCGGAGGC 2689
QY 1730 CAGGAACTGAGTGAAGCCCTGCTGCTGCTGCTCAGGATGGGTCTGGGGCCCTCGTGT 1789
Db 2690 CAGGAACTGAGTGAAGCCCTGCTGCTGCTGCTCAGGATGGGTCTGGGGCCCTCGTGT 2749
QY 1790 CATCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAAACCCAGGACTT 1849
Db 2750 CATCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAAACCCAGGACTT 2809
QY 1850 GGCTCCCCATCTCAGAACTCACAAGCCATGCTCCCCAGCTGGGAACCTCAACCTCC 1909
Db 2810 GGCTCCCCATCTCAGAACTCACAAGCCATGCTCCCCAGCTGGGAACCTCAACCTCC 2869
QY 1910 CCCTCCCTCGGTGGTGCAGAGGGGTGGGACAGGGCGGGGGTTCGCCCTGTACAT 1969
Db 2870 CCCTCCCTCGGTGGTGCAGAGGGGTGGGACAGGGCGGGGGTTCGCCCTGTACAT 2929
QY 1970 ACCCTGCCATACCAACCCAGGTATTAATCTCGTGGTGTGTTTTTATTTTATTTT 2029
Db 2930 ACCCTGCCATACCAACCCAGGTATTAATCTCGTGGTGTGTTTTTATTTTATTTT 2989
QY 2030 TTGTTTTGATTTTTTAAATAAGAAATTTTCATTTTAAGCACAATAAAA 2076
Db 2990 TTGTTTTGATTTTTTAAATAAGAAATTTTCATTTTAAGCACAATAAAA 3036

RESULT 4
AAV64990
ID AAV64990 standard; cDNA; 3036 BP.
XX
AC AAV64990;
XX
AC AAV64990;
DT 05-FEB-1999 (first entry)
XX
DE Human myl/RAR-alpha fusion protein cDNA.
XX
KW Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;
KW acute promyelocytic leukaemia; APL; t(15;17); translocation;
KW treatment; all-trans retinoic acid; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 67..2460
FT /*tag= a
FT /product= "myl/RAR-alpha fusion protein"
XX
PN US5843642-A.
XX
PD 01-DEC-1998.
XX
PF 21-JUL-1993; 93US-0095728.
XX
PR 21-JUL-1993; 93US-0095728.
PR 22-MAR-1991; 91US-0673838.
PR 22-MAR-1991; 91US-0675084.
XX
PA (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI Dmitrovsky E, Frankel S, Miller WH, Warrell RP;
XX
XX WPI; 1999-044563/04.
DR P-PSDB; AAW81963.
XX
PT Diagnosis of acute promyelocytic leukaemia - by detecting nucleic
PT acid encoding abnormal retinoic acid receptor-alpha
XX
PS Disclosure; Fig 4A-G; 38pp; English.
XX
```

CC This sequence encodes a human myl/RAR-alpha (retinoic acid receptor
CC alpha) fusion protein which is used in a method for identifying a
CC subject with acute promyelocytic leukaemia (APL) resulting from a
CC t(15;17) translocation who will respond to treatment with all-trans
CC retinoic acid. The protein can also be used to identify a subject
CC with indications of APL who will not respond to treatment with all-trans
CC retinoic acid.

XX
SQ Sequence 3036 BP; 605 A; 1018 C; 890 G; 523 T; 0 other;

Query Match 85.48; Score 1782.2; DB 20; Length 3036;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 290 CCATTGAGACCCAGAGCAGCAGTCTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACCCC 349
Db 1250 CCATTGAGACCCAGAGCAGCAGTCTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACCCC 1309
QY 350 CTCTACCCCGCATCTACAAGCCTTGTCTGTCTGAGGACAAGTCTCAGGCTACCACT 409
Db 1310 CTCTACCCCGCATCTACAAGCCTTGTCTGTCTGAGGACAAGTCTCAGGCTACCACT 1369
QY 410 ATGGGGTTCAGGCGCTGTGAGGGCTGTCAAGGGCTTCTTCCGCCGAGCATCCAGAAGAACA 469
Db 1370 ATGGGGTTCAGGCGCTGTGAGGGCTGTCAAGGGCTTCTTCCGCCGAGCATCCAGAAGAACA 1429
QY 470 TGTGTACAGTGTACCCGGGACAGAACTGCATCATCAACAAGGTGACCCGGGAACCCCT 529
Db 1430 TGTGTACAGTGTACCCGGGACAGAACTGCATCATCAACAAGGTGACCCGGGAACCCCT 1489
QY 530 GCCAGTACTGCCAGTCTCAGAAAGTCTTGAAGTGGGCATGTCCAAGAGTCTGTGAGAA 589
Db 1490 GCCAGTACTGCCAGTCTCAGAAAGTCTTGAAGTGGGCATGTCCAAGAGTCTGTGAGAA 1549
QY 590 ACGACGAAACAAGAAAGAGAGTGCCCAAGCCCGAGTGTCTGAGAGCTACACGC 649
Db 1550 ACGACGAAACAAGAAAGAGAGTGCCCAAGCCCGAGTGTCTGAGAGCTACACGC 1609
QY 650 TGACGCGGAGGTGGGGAGCTCATTTAGAGAGTGGCCAAAGCCGACCAAGAAACCTTCC 709
Db 1610 TGACGCGGAGGTGGGGAGCTCATTTAGAGAGTGGCCAAAGCCGACCAAGAAACCTTCC 1669
QY 710 CTGCCCTCTGCCAGTCTGGGCAATATACACTACGAACAACAGCTCAGAAACAAGTCTCTC 769
Db 1670 CTGCCCTCTGCCAGTCTGGGCAATATACACTACGAACAACAGCTCAGAAACAAGTCTCTC 1729
QY 770 TGGACATTGACCTCTGGGACAAGTTCAGTGAAGTCTCCACCAAGTGTGATTAAGACTG 829
Db 1730 TGGACATTGACCTCTGGGACAAGTTCAGTGAAGTCTCCACCAAGTGTGATTAAGACTG 1789
QY 830 TGGAGTTCGCCAAGCAGCTGCCGGCTTACACAGCTTACCATTCGCCGACCAAGTACCC 889
Db 1790 TGGAGTTCGCCAAGCAGCTGCCGGCTTACACAGCTTACCATTCGCCGACCAAGTACCC 1849
QY 890 TCCTCAAGGTGTGCTGTGCATCTGATCTGCGGATCTGCACGCGGTACACGCCCG 949
Db 1850 TCCTCAAGGTGTGCTGTGCATCTGATCTGCGGATCTGCACGCGGTACACGCCCG 1909
QY 950 AGCAGACACCATGACCTTCTCGGAGGGGTGACCTTGAAACCGGACCCAGATGCACAAG 1009
Db 1910 AGCAGACACCATGACCTTCTCGGAGGGGTGACCTTGAAACCGGACCCAGATGCACAAG 1969
QY 1010 CTGGCTTCGGCCCTTCACCGAGCTGGTCTTTCGCTTCGCCCAACCAAGCTGCTGCCCTGG 1069
Db 1970 CTGGCTTCGGCCCTTCACCGAGCTGGTCTTTCGCTTCGCCCAACCAAGCTGCTGCCCTGG 2029
QY 1070 AGATGATGATCGGAGAGCGGGCTGCTCAGCGCCATCTGCCCTCATCTGCGGAGACGCC 1129
Db 2030 AGATGATGATCGGAGAGCGGGCTGCTCAGCGCCATCTGCCCTCATCTGCGGAGACGCC 2089
QY 1130 AGGACTGGAGCAGCGGACCGGGTGGACATGCTGAGGAGCGGCTGCTGGAGGCGCTAA 1189
Db 2090 AGGACTGGAGCAGCGGACCGGGTGGACATGCTGAGGAGCGGCTGCTGGAGGCGCTAA 2149
```

```
QY 1190 AGGTCTAGCTGCGGAGCGAGCGCCAGCCGCCACATGTTCCCAAGATGCTAATGA 1249
Db 2150 AGGTCTAGCTGCGGAGCGAGCGCCAGCCGCCACATGTTCCCAAGATGCTAATGA 2209
QY 1250 AGATTACTGACCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGTGATCACGCTGAAGA 1309
Db 2210 AGATTACTGACCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGTGATCACGCTGAAGA 2269
QY 1310 TGGAGATCCCGGGCTCCATGCCCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGGCC 1369
Db 2270 TGGAGATCCCGGGCTCCATGCCCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGGCC 2329
QY 1370 TGGACACTCTGAGCGGACAGCCGGGGGTGGGGGGGACGGGGTGGCTGCCGCCCC 1429
Db 2330 TGGACACTCTGAGCGGACAGCCGGGGGTGGGGGGGACGGGGTGGCTGCCGCCCC 2389
QY 1430 CGCCAGGAGCTGTAGCCCGCCAGCCCTCAGCCCCAGCTCCCAACAGAGACGCGCCGACCC 1489
Db 2390 CGCCAGGAGCTGTAGCCCGCCAGCCCTCAGCCCCAGCTCCCAACAGAGACGCGCCGACCC 2449
QY 1490 ACTCCCGCTGACCGCCCGCCAGCCCATGACACAGAGCCCTCGCCCTCGCCCGCTTTTCT 1549
Db 2450 ACTCCCGCTGACCGCCCGCCAGCCCATGACACAGAGCCCTCGCCCTCGCCCGCTTTTCT 2509
QY 1550 CTGCTTTTCTACCGACCATGTGACCCCGGACAGCCCTCGCCCGCTCGCCCGCTCGCGGC 1609
Db 2510 CTGCTTTTCTACCGACCATGTGACCCCGGACAGCCCTCGCCCGCTCGCCCGCTCGCGGC 2569
QY 1610 AGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1669
Db 2570 AGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2629
QY 1670 GCCTGGGCGCTCAGTGGAGTGCCTCTCCACAGAGCTGGGCTCAGCTCAGAGGCCGAGGC 1729
Db 2630 GCCTGGGCGCTCAGTGGAGTGCCTCTCCACAGAGCTGGGCTCAGCTCAGAGGCCGAGGC 2689
QY 1730 CAGGAAGTGTGAGGCGCCCTGGTCTGGTCTCAGGATGGGTCTGGGGGCTCGTGT 1789
Db 2690 CAGGAAGTGTGAGGCGCCCTGGTCTGGTCTCAGGATGGGTCTGGGGGCTCGTGT 2749
QY 1790 CATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGCGGAGCTT 1849
Db 2750 CATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGCGGAGCTT 2809
QY 1850 GGCTCCCGCCATCTCAGAACTCAGAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 1909
Db 2810 GGCTCCCGCCATCTCAGAACTCAGAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 2869
QY 1910 CCCCTGCTCGGTTGGTGACAGAGGGGTGGGACAGGGCGGGGGTTCGCCCTCTACAT 1969
Db 2870 CCCCTGCTCGGTTGGTGACAGAGGGGTGGGACAGGGCGGGGGTTCGCCCTCTACAT 2929
QY 1970 ACCCTGCATACCAACCCAGGTATTAATCTCGTGGTTTGTGTTTATTTATTTT 2029
Db 2930 ACCCTGCATACCAACCCAGGTATTAATCTCGTGGTTTGTGTTTATTTATTTT 2989
QY 2030 TTGTTTGTATTTTATTAAGAAATTTTCAATTTTAAGCAAAAAA 2076
Db 2990 TTGTTTGTATTTTATTAAGAAATTTTCAATTTTAAGCAAAAAA 3036
```

RESULT 5
AAA38656
ID AAA38656 standard; DNA; 3511 BP.
XX
AC AAA38656;
AC
XX
DT 11-SEP-2000 (first entry)
XX
DE Human PML/alpha-RAR fusion gene, SEQ ID NO:13.
XX
KW External guide sequence; EGS; RNase P; human; ribozyme;

KW catalytic RNA: antiviral; hepatitis B virus; HBV; anticancer; APL;
KW acute promyelocytic leukaemia; retinoic acid receptor-alpha;
KW RAR-alpha; PML; fusion gene; ss.
XX Homo sapiens.
XX US6057153-A.
XX 02-MAY-2000.
XX 14-JUL-1997; 97US-0892747.
XX 13-JAN-1995; 95US-0372556.
XX 19-JAN-1996; 96WO-US00513.
XX (UYUA) UNIV YALE.
XX Pace U, Goldberg AR, Werner M, George ST, Ma M;
XX WPI; 2000-399010/34.
XX Novel modified external guide sequences for eukaryotic RNase P for
PT cleaving hepatitis viral RNA comprises RNase P targeting sequence and
PT recognition sequence complementary to targeted sequence in target RNA
PT
XX
XX Example 1; Column 45-48; 76pp; English.
XX The invention relates to modified human RNase P external guide sequences
CC (EGSSs). The EGS is a ribozyme which forms the catalytic domain of the
CC ribonucleoprotein RNase P, which is involved in the processing of tRNA.
CC The EGS comprises a cleavage targeting sequence and a recognition
CC sequence which is complementary to the target RNA sequence. The
CC recognition sequence comprises an A and a D recognition arm which
CC correspond to the aminoacyl acceptor stem and the D stem of tRNA. The A
CC recognition arm is located at the 3' end of the EGS and the D at the 5'
CC end. The cleavage targeting sequence forms a structure corresponding to
CC the T stem and T loop of precursor tRNA. The modified EGSs of the
CC invention comprise at least one modified nucleotide or an unmodified
CC deoxyribonucleotide, and the 3' end of the molecule is preferably
CC modified. The A and D arms, and the T stem and loop can consist of
CC ribonucleotides with 5'-phosphate or 5'-phosphorothioate linkages;
CC 2'-modified ribonucleotides with 5'-phosphate or 5'-phosphorothioate
CC linkages; or deoxyribonucleotides with 5'-phosphate or 5'-
CC phosphorothioate linkages. EGSs of the invention are used for cleaving
CC a target RNA molecule. The invention discloses EGSs which are targeted
CC to hepatitis B virus (HBV) RNA sequences, particularly hepatitis B
CC surface antigen (HBsAg) RNA, and EGSs targeted at PML/retinoic acid
CC receptor alpha (RAR-alpha) fusion gene mRNA which is found in the
CC majority of acute promyelocytic leukaemia (APL) cases. Sequences
CC AAA38652- AAA38655 represent active and control PML/RAR-alpha fusion gene
CC RNA- targeted EGSs used in an exemplification of the invention. The
CC target RNA (AAA38651) comprises the transcript of the junction between
CC the PML and the RAR-alpha gene fragments of the fusion gene. The fusion
CC gene is generated by a balanced reciprocal chromosomal translocation
CC between the long arms of chromosome 15 and 17, and the fusion protein
CC encoded is a dominant negative regulator of myeloid differentiation.
XX
XX Sequence 3511 BP; 715 A; 1168 C; 1038 G; 589 T; 1 other;
Query Match 85.4%; Score 1782.2; DB 21; Length 3511;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 290 CCATTGAGACCCAGCAGCAGTCTGAGAGATAGTCCCGAGCCCTCCCTCGCCACCCC 349
DB 1725 CCATTGAGACCCAGCAGCAGTCTGAGAGATAGTCCCGAGCCCTCCCTCGCCACCCC 1784
QY 350 CTCCTACCCCGCATCTACAAAGCCTTGTCTGTCAGAGCAAGTCCCTCAGGCTACCACT 409
DB 1785 CTCCTACCCCGCATCTACAAAGCCTTGTCTGTCAGAGCAAGTCCCTCAGGCTACCACT 1844
QY 410 ATGGGGTGCAGCGCCCTGTGAGGGCTGCAAGGGCTTCTTCCGGCGCAGCATCCAGAGAACA 469

DB 1845 ATGGGGTGCAGCGCCCTGTGAGGGCTGCAAGGGCTTCTTCCGCCCGCAGCATCCAGAGAACA 1904
QY 470 TGGTGTACACAGTGTACACGGGACAAAGAACTGCATCATCAACAGGTGACCCGGAACCCCT 529
DB 1905 TGGTGTACACAGTGTACACGGGACAAAGAACTGCATCATCAACAGGTGACCCGGAACCCCT 1964
QY 530 GCCAGTACTGCCGACTGCGAGAAGTCTTTGAAAGTGGGCATGTCCAAGGAGTCTGTGAGAA 589
DB 1965 GCCAGTACTGCCGACTGCGAGAAGTCTTTGAAAGTGGGCATGTCCAAGGAGTCTGTGAGAA 2024
QY 590 ACGACCGAAACAAGAAAGAGAGGTGCCCAAGCCGAGTCTCTGAGAGCTACACGC 649
DB 2025 ACGACCGAAACAAGAAAGAGAGGTGCCCAAGCCGAGTCTCTGAGAGCTACACGC 2084
QY 650 TGACGCCGAGGTGGGGAGTCTATTGAGAAGTGGCAAGGCGCAAGGCGCACAGAAACCTTCC 709
DB 2085 TGACGCCGAGGTGGGGAGTCTATTGAGAAGTGGCAAGGCGCAAGGCGCACAGAAACCTTCC 2144
QY 710 CTGCCCCTCTGCCAGCTGGGCAAAATACACTACGNAACAACAGCTCAGAACACAGTGTCTCTC 769
DB 2145 CTGCCCCTCTGCCAGCTGGGCAAAATACACTACGNAACAACAGCTCAGAACACAGTGTCTCTC 2204
QY 770 TGGACATTGACCTCTGGGACAAAGTTTCAGTGAACCTCTCCACCAAGTGCATCATTAAGACTG 829
DB 2205 TGGACATTGACCTCTGGGACAAAGTTTCAGTGAACCTCTCCACCAAGTGCATCATTAAGACTG 2264
QY 830 TGGAGTTGCCCAAGCAGCTGCCCGGCTTCACCACCCCTCACCATCGCGGACAGATCACCC 889
DB 2265 TGGAGTTGCCCAAGCAGCTGCCCGGCTTCACCACCCCTCACCATCGCGGACAGATCACCC 2324
QY 890 TCCTCAAGCTGCTGCTGGACATCTCTGATCTCGGGATCTGCACGCGGTACACGCCCG 949
DB 2325 TCCTCAAGCTGCTGCTGGACATCTCTGATCTCGGGATCTGCACGCGGTACACGCCCG 2384
QY 950 ACGAGACACCATGACCTTCTCGGAGGGTGTGACCTGAAACCGGACCCAGATGCACAACG 1009
DB 2385 ACGAGACACCATGACCTTCTCGGAGGGTGTGACCTGAAACCGGACCCAGATGCACAACG 2444
QY 1010 CTGGCTTGGGCCCCCTACCGACCTGGTCTTTGGCTTGGCCAAACAGCTGCTGCCCTCG 1069
DB 2445 CTGGCTTGGGCCCCCTACCGACCTGGTCTTTGGCTTGGCCAAACAGCTGCTGCCCTCG 2504
QY 1070 AGATGATGATCGGGAGAGGGGCTCTCAGCGCCATCTGCCTCATCTGCGGAGACGCC 1129
DB 2505 AGATGATGATCGGGAGAGGGGCTCTCAGCGCCATCTGCCTCATCTGCGGAGACGCC 2564
QY 1130 AGSACTGAGCAGCCGGACCGGGTGGACATGCTGCAGGAGCGCTGCTGGAGGCGCTAA 1189
DB 2565 AGSACTGAGCAGCCGGACCGGGTGGACATGCTGCAGGAGCGCTGCTGGAGGCGCTAA 2624
QY 1190 AGSCTTACGTGGGAGAGCGAGCGCCCGCCCGCCACATGTTCCCAAGATGCTAATGA 1249
DB 2625 AGSCTTACGTGGGAGAGCGAGCGCCCGCCCGCCACATGTTCCCAAGATGCTAATGA 2684
QY 1250 AGATTACTGACCTGGAAGCATCAGCGCCCAAGGGGCTGAGCGGGTGTATCAGCTGAAGA 1309
DB 2685 AGATTACTGACCTGGAAGCATCAGCGCCCAAGGGGCTGAGCGGGTGTATCAGCTGAAGA 2744
QY 1310 TGGAGATCCCGGGCTCCATCGCCCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGGCC 1369
DB 2745 TGGAGATCCCGGGCTCCATCGCCCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGGCC 2804
QY 1370 TGGAGACTCTGAGCGGACAGCGGGGGTGGGGGGCGGAGCGGGGTGGCTGCGCCCGCC 1429
DB 2805 TGGAGACTCTGAGCGGACAGCGGGGGTGGGGGGCGGAGCGGGGTGGCTGCGCCCGCC 2864
QY 1430 CGCCAGGCAAGCTGTAGCCCCCAGCCTCAGCCCCCAGCTCCAAAGAGCAGCCGCGCCACCC 1489
DB 2865 CGCCAGGCAAGCTGTAGCCCCCAGCCTCAGCCCCCAGCTCCAAAGAGCAGCCGCGCCACCC 2924
QY 1490 ACTCCCGTGCAGCGCCCGCCAGCCACATGGACACAGCCCTCGCCCTCGCCCGCGCTTTCT 1549

```
Db 2925 ACTCCCGGTGACCGCCACGCCACATGACACAGCCCTCGCCCTCCCGCGGCTTTCT 2984
QY 1550 CTGCTTTCTACGACCATGTGACCCCGCAGCCCTGCCCCACCTGCTCCCTCCCGGC 1609
Db 2985 CTGCTTTCTACGACCATGTGACCCCGCAGCCCTGCCCCACCTGCTCCCTCCCGGC 3044
QY 1610 AGTACTGGGACCTTCCCTGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1669
Db 3045 AGTACTGGGACCTTCCCTGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3104
QY 1670 GCTTGGGCTCTAGTGGAGCTGCTGCTCCACAGCTGCTGCTGCTGCTGCTGCTG 1729
Db 3105 GCTTGGGCTCTAGTGGAGCTGCTGCTCCACAGCTGCTGCTGCTGCTGCTGCTG 3164
QY 1730 CAGGAAGTCTAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1789
Db 3165 CAGGAAGTCTAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3224
QY 1790 CATCAAGACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1849
Db 3225 CATCAAGACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3284
QY 1850 GCTTCCCTCATCTCAGAACTCAGAGCCATGCTGCTGCTGCTGCTGCTGCTGCTG 1909
Db 3285 GCTTCCCTCATCTCAGAACTCAGAGCCATGCTGCTGCTGCTGCTGCTGCTGCTG 3344
QY 1910 CCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1969
Db 3345 CCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3404
QY 1970 ACCCTGCCATACAAACCCAGGTATTAATCTCGCTGCTGCTGCTGCTGCTGCTG 2029
Db 3405 ACCCTGCCATACAAACCCAGGTATTAATCTCGCTGCTGCTGCTGCTGCTGCTG 3464
QY 2030 TGTGTTTGAATTTTAAAGAAATTTTCAATTTTAAAGCAAAAAA 2076
Db 3465 TGTGTTTGAATTTTAAAGAAATTTTCAATTTTAAAGCAAAAAA 3511

RESULT 6
AAT33259
ID AAT33259 standard; cDNA; 3511 BP.
XX
AC AAT33259;
XX
DT 09-OCT-1996 (first entry)
XX
DE PML-RAR-alpha DNA sequence.
XX
KW Hammerhead ribozyme; acute promyelocytic leukaemia; APL;
KW lymphoma; therapy; PML-RAR-alpha; retinoic acid receptor;
KW external guide sequence; EGS; antisense; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1..1724
FT /*tag= a
FT /note= "PML DNA"
FT misc_difference 1477
FT /*tag= b
FT /note= "base n at position 1477 is not identified
FT /note= in the specification"
FT misc_feature 1725..3511
FT /*tag= c
FT /note= "RAR-alpha DNA"
XX
PN WO9618733-A2.
XX
PD 20-JUN-1996.
XX
PF 14-DEC-1995; 95WO-US16451.
XX
```

```
PR 14-DEC-1994; 94US-0354956.
XX
PA (INNO-) INNOVIR LAB INC.
XX
PI George ST, Goldberg AR, Pace U;
XX
DR WPI; 1996-300650/30.
XX
PT RNA construct(s) including ribozyme(s) and antisense
PT oligo:nucleotide(s) - for the inactivation of RNA associated with,
PT e.g. promyelocytic leukaemia or follicular lymphoma
XX
PS Example 1; Page 50-52; 81pp; English.
XX
CC PML-RAR-alpha DNA (AAT33259) is associated with acute promyelocytic
CC leukemia (APL). APL is characterised by a balanced, reciprocal
CC translocation between the long arms of chromosomes 15 and 17,
CC resulting in a fusion of the retinoic acid receptor gene (RAR-alpha)
CC and a gene for a putative transcription factor, PML. Methods for
CC treating APL involve the use of ribozymes, external guide sequences
CC and antisense oligonucleotides (see also AAT33245 and AAT33251) that
CC specifically cleave the PML-RAR-alpha fusion mRNA (see also AAT33251)
CC but not wild-type RAR-alpha mRNA. Plasmids carrying portions of the
CC PML-RAR-alpha and RAR-alpha genes (see also AAT33260), allow synthesis
CC of shortened versions of APL mRNA in vitro, facilitating testing and
CC screening processes.
XX
SQ Sequence 3511 BP; 715 A; 1168 C; 1038 G; 589 T; 1 other;

Query Match 85.3%; Score 1779; DB 17; Length 3511;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1782; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 290 CCATTGAGACCCAGACAGCAGTCTCTGAAGAGATAGTGCCAGCCCTCCCTCGCACCCC 349
Db 1725 CCATTGAGACCCAGACAGCAGTCTCTGAAGAGATAGTGCCAGCCCTCCCTCGCACCCC 1784
QY 350 CTCCTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAAGTCTCAGGTACCACT 409
Db 1785 CTCCTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAAGTCTCAGGTACCACT 1844
QY 410 ATGGGGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTCCGCCCGCAGCATCCAGAAACA 469
Db 1845 ATGGGGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTCCGCCCGCAGCATCCAGAAACA 1904
QY 470 TGGTGTACACGTGTACACCGGGACAAAGTGTGATCATCATCAACAAGGTGACCCGAACCCCT 529
Db 1905 TGGTGTACACGTGTACACCGGGACAAAGTGTGATCATCATCAACAAGGTGACCCGAACCCCT 1964
QY 530 GCCAGTACTGCCGACTGCGAGAGTGTCTTTGAAGTGGGCGATGTCCAAAGGAGTCTGTGAGAA 589
Db 1965 GCCAGTACTGCCGACTGCGAGAGTGTCTTTGAAGTGGGCGATGTCCAAAGGAGTCTGTGAGAA 2024
QY 590 ACGACCGAAACAAGAAAGAGGAGGTGCCCAAGCCCGAGTGTCTCTGAGAGTACACGC 649
Db 2025 ACGACCGAAACAAGAAAGAGGAGGTGCCCAAGCCCGAGTGTCTCTGAGAGTACACGC 2084
QY 650 TGACCCCGGAGTGGGGAGCTCATTTGAGAAGGTGGCGAAAGCGCACCAAGAAACCTTCC 709
Db 2085 TGACCCCGGAGTGGGGAGCTCATTTGAGAAGGTGGCGAAAGCGCACCAAGAAACCTTCC 2144
QY 710 CTGCCCCCTGCCAGCTGGGCAATATACACTACGAAACAGCTCAGAAACACGCTGTCTCTC 769
Db 2145 CTGCCCCCTGCCAGCTGGGCAATATACACTACGAAACAGCTCAGAAACACGCTGTCTCTC 2204
QY 770 TGGACATTGACCTCTGGGACAAAGTTCAGTGAAGTCTCCACCAAGTGCATCATTAAGACTG 829
Db 2205 TGGACATTGACCTCTGGGACAAAGTTCAGTGAAGTCTCCACCAAGTGCATCATTAAGACTG 2264
QY 830 TGGAGTTCGCCAAGCAGCTGCCCGGCTTCAACCAACCTCACCAGTGCAGACAGTACACCC 889
Db 2265 TGGAGTTCGCCAAGCAGCTGCCCGGCTTCAACCAACCTCACCAGTGCAGACAGTACACCC 2324
```

QY 890 TCCTCAAGGCTGCTGCTGACATCTGATCTCGGAGTCTGCACGCGGTACACGCCG 949
Db 2325 TCCTCAAGGCTGCTGCTGACATCTGATCTCGGAGTCTGCACGCGGTACACGCCG 2384
QY 950 AGCAGACACCATGACCTTCTCGGAGGCTGACCTTGAACCGGACCCAGATGCACAACG 1009
Db 2385 AGCAGACACCATGACCTTCTCGGAGGCTGACCTTGAACCGGACCCAGATGCACAACG 2444
QY 1010 CTGGCTTCGGCCCTCCCTACCGACCTGCTCTTGGCTTCGCTCCCAACCAAGCTGCTGCCCTGG 1069
Db 2445 CTGGCTTCGGCCCTCCCTACCGACCTGCTCTTGGCTTCGCTCCCAACCAAGCTGCTGCCCTGG 2504
QY 1070 AGATGATGATGCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGACCGCC 1129
Db 2505 AGATGATGATGCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGACCGCC 2564
QY 1130 AGACCTTGAGCAGCCGCGGAGCGGCTGATGCTGAGGAGCGCTGCTGGAGCGCTAA 1189
Db 2565 AGACCTTGAGCAGCCGCGGAGCGGCTGATGCTGAGGAGCGCTGCTGGAGCGCTAA 2624
QY 1190 AGTCTACCTGCGGAGCGGAGCGGCTGATGCTGAGGAGCGCTGCTGGAGCGCTAA 1249
Db 2625 AGTCTACCTGCGGAGCGGAGCGGCTGATGCTGAGGAGCGCTGCTGGAGCGCTAA 2684
QY 1250 AGATTACTGACCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGCTGATCAGCTGAAGA 1309
Db 2685 AGATTACTGACCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGCTGATCAGCTGAAGA 2744
QY 1310 TGGAGATCCGGGCTCCATCGCCCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC 1369
Db 2745 TGGAGATCCGGGCTCCATCGCCCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC 2804
QY 1370 TGGACACTTGAGCGGACAGCGGGGGTGGGGGGCGGACGGGGTGGCTGCCCTCCGCC 1429
Db 2805 TGGACACTTGAGCGGACAGCGGGGGTGGGGGGCGGACGGGGTGGCTGCCCTCCGCC 2864
QY 1430 CGCCAGGAGCTGTAGCCCGACGCTCAGCGCCAGCTCCACAGAAAGCGCGGCGCACCC 1489
Db 2865 CGCCAGGAGCTGTAGCCCGACGCTCAGCGCCAGCTCCACAGAAAGCGCGGCGCACCC 2924
QY 1490 ACTCCCGTGGCCCGCCAGCCACATGAGACAGCCCTCGCCCTCCCGCCCGCTTTTCT 1549
Db 2925 ACTCCCGTGGCCCGCCAGCCACATGAGACAGCCCTCGCCCTCCCGCCCGCTTTTCT 2984
QY 1550 CTGCTTTCTACCGACCATGTGACCCCGCAGCGCCCTCCCGCCCTCCCGCCCGG 1609
Db 2985 CTGCTTTCTACCGACCATGTGACCCCGCAGCGCCCTCCCGCCCTCCCGCCCGG 3044
QY 1610 AGTACTGGGACCTTCCCTGGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669
Db 3045 AGTACTGGGACCTTCCCTGGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3104
QY 1670 GCCTGGGCGCTCAGTGGAGTGTCTGCTCCACAGCTGGGCTGACCTCAGAGGCGGAGG 1729
Db 3105 GCCTGGGCGCTCAGTGGAGTGTCTGCTCCACAGCTGGGCTGACCTCAGAGGCGGAGG 3164
QY 1730 CAGGAACCTAGTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1789
Db 3165 CAGGAACCTAGTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3224
QY 1790 CATCAAGACACCCCTTCCCGAGCTCACCACATCTCATCAGCAAGCAAGCGGAGCTT 1849
Db 3225 CATCAAGACACCCCTTCCCGAGCTCACCACATCTCATCAGCAAGCAAGCGGAGCTT 3284
QY 1850 GGTCTCCCGCTCCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 1909
Db 3285 GGTCTCCCGCTCCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 3344
QY 1910 CCCCTGCTCGGTTGGTGACAGAGGGGTGGGACAGGGGGGGGTTCCCTCTGTACAT 1969
Db 3345 CCCCTGCTCGGTTGGTGACAGAGGGGTGGGACAGGGGGGGGTTCCCTCTGTACAT 3404
QY 1970 ACCCTGCCATACCAACCCAGGTATTAATTCGCTGCTGCTGCTGCTGCTGCTGCTG 2029

Db 3405 ACCCTGCCATACCAACCCAGGTATTAATTCGCTGCTGCTGCTGCTGCTGCTGCTG 3464
QY 2030 TTGTTTGTGATTTTAAATAAGATTTTCATTTTAAAGCACAAAAA 2076
Db 3465 TGTTTGTGATTTTAAATAAGATTTTCATTTTAAAGCACAAAAA 3511
RESULT 7
ABK84517
ID ABK84517 standard; cDNA; 2907 BP.
XX AC ABK84517;
XX DT 14-AUG-2002 (first entry)
XX Human cDNA differentially expressed in granulocytic cells #1088.
DE Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX Homo sapiens.
XX WO200228999-A2.
PN 11-APR-2002.
XX 03-OCT-2001; 2001WO-US30821.
PF -03-OCT-2000; 2000US-237189P.
PR (GENE-) GENE LOGIC INC.
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
WPI; 2002-435328/46.
PT Detecting granulocyte activation by detecting differential expression
of genes associated with granulocyte activation, which serves as
diagnostic markers that is useful for monitoring disease states and
drug toxicity
XX Claim 1; SEQ ID No 1088; 114pp; English.
PS The invention relates to detecting (M1) granulocyte (GC) activation
(GCA), by detecting the level of expression of gene(s) (Gs) identified by
DNA chip analysis as given in the specification, and comparing
the expression level to an expression level in an unactivated
GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
that alters the expression of at least one gene in Gs; (2) screening (M3)
for an agent capable of modulating GCA or an inflammation (especially
chronic) in a tissue, an allergic response in a subject, exposure of a
subject to a pathogen or sterile inflammatory disease using the
gene expression profile; (3) detecting (M4) an inflammation (especially
chronic) in a tissue, an allergic response in a subject, exposure of a
subject to a pathogen or sterile inflammatory disease, by detecting the
level of expression in a sample of the tissue of gene(s) from Gs, where
the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
an allergic response in a subject, exposure of a subject to a pathogen
or sterile inflammatory disease, by contacting a tissue having
inflammation with an agent that modulates the expression of gene(s)
from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
modulating GA; M3 is useful for screening an agent capable of modulating
GCA preferably in an inflammation in a tissue; M4 is useful for
detecting an inflammation (especially chronic) in a tissue, an allergic

CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC peridontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ

Sequence 2907 BP; 561 A; 1011 C; 789 G; 546 T; 0 other;

Query Match 85.3%; Score 1778.8; DB 24; Length 2907;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1786; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 288 CTCATTGAGACCCAGAGCAGCAGTCTTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC 347
DB 279 CACCATGAGACCCAGAGCAGCAGTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC 338
QY 348 CCCTCTACCCCGCATCTACAAAGCCTTCTGTCTGTGAGGACAGTCCCTCAGGCTACCA 407
DB 339 CCCTCTACCCCGCATCTACAAAGCCTTCTGTCTGTGAGGACAGTCCCTCAGGCTACCA 398
QY 408 CTATGGGGTCAGCGCCTGTGAGGCTGCAAGGGCTTCTCCGCGAGCATCCAGAGAA 467
DB 399 CTATGGGGTCAGCGCCTGTGAGGCTGCAAGGGCTTCTCCGCGAGCATCCAGAGAA 458
QY 468 CATGGGTACAGTGTACCGGGACAGAGTGTATCATCAAGGTGACCGGACCC 527
DB 459 CATGGGTACAGTGTACCGGGACAGAGTGTATCATCAAGGTGACCGGACCC 518
QY 528 CTCGAGTACTCGCGACTGCAGAAAGTCTTGAAGTGGGATGTCACAGGAGTCTGTGAG 587
DB 519 CTCGAGTACTCGCGACTGCAGAAAGTCTTGAAGTGGGATGTCACAGGAGTCTGTGAG 578
QY 588 AACAGCCGAAACAGAAAGAGAGTGTCCCAAGCCGAGTGTCTGTGAGAGTACAC 647
DB 579 AACAGCCGAAACAGAAAGAGAGTGTCCCAAGCCGAGTGTCTGTGAGAGTACAC 638
QY 648 GCTGACCCGAGTGGGGAGTCTATTGAGAGTGTCCCAAGCCGAGTGTCTGTGAGAGTAC 707
DB 639 GCTGACCCGAGTGGGGAGTCTATTGAGAGTGTCCCAAGCCGAGTGTCTGTGAGAGTAC 698
QY 708 CCCTGCCCTCTGCCAGTGGGCAATACACTACGAAACAGAGTCTCAGAAACACGCTGTCTC 767
DB 699 CCCTGCCCTCTGCCAGTGGGCAATACACTACGAAACAGAGTCTCAGAAACACGCTGTCTC 758
QY 768 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAATCTCCACCAAGTGCATCATTAAGAC 827
DB 759 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAATCTCCACCAAGTGCATCATTAAGAC 818
QY 828 TGTGGAGTTCGCAAGCAGTGTCCCGGGTTCACCAACCTCACCATCGCGACAGATCAC 887
DB 819 TGTGGAGTTCGCAAGCAGTGTCCCGGGTTCACCAACCTCACCATCGCGACAGATCAC 878
QY 888 CCTCCTCAAGGCTGCCCTGGACATCTGTATCTCGGATCTGACGGGTACAGGCC 947
DB 879 CCTCCTCAAGGCTGCCCTGGACATCTGTATCTCGGATCTGACGGGTACAGGCC 938
QY 948 CGAGCAGGACACCATGACCTTCTCGGACGGGCTGACCTCAACCGGACCCAGATGCACAA 1007
DB 939 CGAGCAGGACACCATGACCTTCTCGGACGGGCTGACCTCAACCGGACCCAGATGCACAA 998
QY 1008 CGCTGGCTTCGGCCCCCTCACCAGCTGTGCTTTGGCTTCGCCAACAGCTGTGCCCT 1067
DB 999 CGCTGGCTTCGGCCCCCTCACCAGCTGTGCTTTGGCTTCGCCAACAGCTGTGCCCT 1058
QY 1068 GGAGATGGATGATGCGGAGAGCGGGGTGCTCAGCGCCATCTGCCTCATCTGCGGAGACCG 1127

DB 1059 GGAGATGGATGATGCGGAGAGCGGGCTGCTCAGCGCCATCTCCCTCATCTGCGGAGACCG 1118
QY 1128 CCAGGACCTGGAGCAGCGGACCGGGTGGACATGTGAGGAGCGGCTGCTGAGAGCGCT 1187
DB 1119 CCAGGACCTGGAGCAGCGGACCGGGTGGACATGTGAGGAGCGGCTGCTGAGAGCGCT 1178
QY 1188 AAGGTCTACGTGGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAT 1247
DB 1179 AAGGTCTACGTGGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAT 1238
QY 1248 GAAGATTACTGACCTGCGAAGCATCAGGCCCAAGGGGCTGAGCGGGTGTATCAGCGTGA 1307
DB 1239 GAAGATTACTGACCTGCGAAGCATCAGGCCCAAGGGGCTGAGCGGGTGTATCAGCGTGA 1298
QY 1308 GATGGAGATCCGGGCTCCATGCGGCTTCTCATCAGGAAATGTTGGAGAACTCAGAGGG 1367
DB 1299 GATGGAGATCCGGGCTCCATGCGGCTTCTCATCAGGAAATGTTGGAGAACTCAGAGGG 1358
QY 1368 CCTGGACACTCTGAGCGACAGCGGGGGTGGGGGCGGAGCGGGTGGGCTGCCCC 1427
DB 1359 CCTGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGAGCGGGTGGGCTGCCCC 1418
QY 1428 CCCGCCAGGACGTGTAGCCCCAGCCTCAGCCCCAGCTCCAACAGAGCAGCCCCGCCAC 1487
DB 1419 CCCGCCAGGACGTGTAGCCCCAGCCTCAGCCCCAGCTCCAACAGAGCAGCCCCGCCAC 1478
QY 1488 CCACCTCCCGTACCGCCCGCCACATGAGACACAGCCCTCGCCCTCGCCCGCCGCTTTT 1547
DB 1479 CCACCTCCCGTACCGCCCGCCACATGAGACAGCCCTCGCCCTCGCCCGCCGCTTTT 1538
QY 1548 CTCCTGCTTTTACCGGACCATGTGACCCCGCAGCAGCCCTGCCCCACCTGCGCTCCCG 1607
DB 1539 CTCCTGCTTTTACCGGACCATGTGACCCCGCAGCAGCCCTGCCCCACCTGCGCTCCCG 1598
QY 1608 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667
DB 1599 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658
QY 1668 AGGCTTGGGCGCTCAGTGGAGTCCCTGCTCCACAGCCCTGGGCTGAGCTCAGAGCGGAG 1727
DB 1659 AGGCTTGGGCGCTCAGTGGAGTCCCTGCTCCACAGCCCTGGGCTGAGCTCAGAGCGGAG 1718
QY 1728 GCCAGGAATCAGTGAAGCCCTTGGTCTGCTGAGTGGGCTCCTGGGGGCTCGTG 1787
DB 1719 GCCAGGAATCAGTGAAGCCCTTGGTCTGCTGAGTGGGCTCCTGGGGGCTCGTG 1778
QY 1788 TTCTATCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAACGCCAGGAC 1847
DB 1779 TTCTATCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAACGCCAGGAC 1838
QY 1848 TTGGCTCCCGCATCTCTAGAACTCACAAGCCATGCTCCCGAGCTGGGGAACCTCAACCT 1907
DB 1839 TTGGCTCCCGCATCTCTAGAACTCACAAGCCATGCTCCCGAGCTGGGGAACCTCAACCT 1898
QY 1908 CCCCCCTGCTCTGCTGTGACAGAGGGGTGGGACAGGGGGGGGTTCCCTCGTAC 1967
DB 1899 CCCCCCTGCTCTGCTGTGACAGAGGGGTGGGACAGGGGGGGTTCCTCGTAC 1958
QY 1968 ATACCTGCCATACCAACCCAGGTATTATTTCTCGCTGGTGTCTTTTATTTTAAATTT 2027
DB 1959 ATACCTGCCATACCAACCCAGGTATTATTTCTCGCTGGTGTCTTTTATTTTAAATTT 2018
QY 2028 TTTTGTGTTGATTTTTTTTAAAGAAATTTTCAATTTAAGCACAAAAAATAAAAAA 2085
DB 2019 TTTTGTGTTGATTTTTTTTAAAGAAATTTTCAATTTAAGCACATTTTACTGAAGAA 2076
RESULT 8
ABL65214
ID ABL65214 standard; DNA; 2907 BP.
XX
AC ABL65214;

XX	15-MAY-2002	(first entry)
DT		
XX		
DE	Lung cancer related gene sequence	SEQ ID NO:3551.
XX		
KW	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;	
KW	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;	
KW	cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;	
XX	gene; ds..	
XX		
OS	Homo sapiens.	
XX		
PN	WO200194529-A2.	
XX		
XX	13-DEC-2001.	
XX		
XX	30-MAY-2001; 2001WO-US10838.	
XX		
PR	05-JUN-2000; 2000US-209473P.	
PR	05-JUN-2000; 2000US-209531P.	
PR	18-SEP-2000; 2000US-233133P.	
PR	18-SEP-2000; 2000US-233617P.	
PR	20-SEP-2000; 2000US-234009P.	
PR	20-SEP-2000; 2000US-234034P.	
PR	20-SEP-2000; 2000US-234052P.	
PR	22-SEP-2000; 2000US-234509P.	
PR	22-SEP-2000; 2000US-234567P.	
PR	25-SEP-2000; 2000US-234923P.	
PR	25-SEP-2000; 2000US-234924P.	
PR	25-SEP-2000; 2000US-235077P.	
PR	25-SEP-2000; 2000US-235082P.	
PR	25-SEP-2000; 2000US-235134P.	
PR	26-SEP-2000; 2000US-235280P.	
PR	26-SEP-2000; 2000US-235637P.	
PR	26-SEP-2000; 2000US-235638P.	
PR	27-SEP-2000; 2000US-235711P.	
PR	27-SEP-2000; 2000US-235720P.	
PR	27-SEP-2000; 2000US-235840P.	
PR	27-SEP-2000; 2000US-235863P.	
PR	28-SEP-2000; 2000US-236028P.	
PR	28-SEP-2000; 2000US-236032P.	
PR	28-SEP-2000; 2000US-236033P.	
PR	28-SEP-2000; 2000US-236034P.	
PR	28-SEP-2000; 2000US-236109P.	
PR	29-SEP-2000; 2000US-236111P.	
PR	29-SEP-2000; 2000US-236842P.	
PR	29-SEP-2000; 2000US-236891P.	
PR	02-OCT-2000; 2000US-237172P.	
PR	02-OCT-2000; 2000US-237173P.	
PR	02-OCT-2000; 2000US-237278P.	
PR	02-OCT-2000; 2000US-237294P.	
PR	02-OCT-2000; 2000US-237295P.	
PR	02-OCT-2000; 2000US-237316P.	
PR	03-OCT-2000; 2000US-237425P.	
PR	03-OCT-2000; 2000US-237598P.	
PR	03-OCT-2000; 2000US-237604P.	
PR	03-OCT-2000; 2000US-237606P.	
PR	03-OCT-2000; 2000US-237608P.	
PR	01-NOV-2000; 2000US-244867P.	
PR	01-NOV-2000; 2000US-245084P.	
XX		
XX	(AVAL-) AVALON PHARM.	
XX		
PI	Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;	
PI	Soppet DR, Weaver Z;	
XX		
XX	WPI; 2002-188264/24.	
XX		
PT	Screening for anti-neoplastic agent involves exposing cells to a	
PT	chemical agent to be tested for anti-neoplastic activity, and	
PT	determining a change in expression of a gene of a signature gene set	
XX		
PS	Claim 1; SEQ ID 3551; 44pp; English.	

```
QY 1008 CGCTGGCTTCCGCCCCCTCACCAGCTGCTGCTTTGGCTTCCGCAACCAAGCTGCTGCCCT 1067
Db 999 CGCTGGCTTCCGCCCCCTCACCAGCTGCTGCTTTGGCTTCCGCAACCAAGCTGCTGCCCT 1058
QY 1068 GGAGATGATGATCGGAGAGCGGGCTGCTCAGGCGCATCTGCCTCATCTCGGAGACCG 1127
Db 1059 GGAGATGATGATCGGAGAGCGGGCTGCTCAGGCGCATCTGCCTCATCTCGGAGACCG 1118
QY 1128 CCAGGACTTGAGAGCGCGGACCGGGTGACATGCTCAGGAGCGCTGCTGGAGCGCT 1187
Db 1119 CCAGGACTTGAGAGCGCGGACCGGGTGACATGCTCAGGAGCGCTGCTGGAGCGCT 1178
QY 1188 AAGGCTACGTCGAGAGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCG 1247
Db 1179 AAGGCTACGTCGAGAGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCG 1238
QY 1248 GAAGATTACTGACCTGCGAAGCATCAGCGCCAGGGGCTGAGCGGTGATCAGCGTAA 1307
Db 1239 GAAGATTACTGACCTGCGAAGCATCAGCGCCAGGGGCTGAGCGGTGATCAGCGTAA 1298
QY 1308 GATGGAGATCCCGGGCTCCATGCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGG 1367
Db 1299 GATGGAGATCCCGGGCTCCATGCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGG 1358
QY 1368 CCTGGACACTCTCAGCGGACAGCGGGGGGTGGGGGCGGAGCGGGTGGCTGCCCTCC 1427
Db 1359 CCTGGACACTCTCAGCGGACAGCGGGGGGTGGGGGCGGAGCGGGTGGCTGCCCTCC 1418
QY 1428 CCGCCAGGCGAGCTGTAGCCCGGAGCTCAGCGCCAGCTTCCAAAGAGAGCGCGGCGAC 1487
Db 1419 CCGCCAGGCGAGCTGTAGCCCGGAGCTCAGCGCCAGCTTCCAAAGAGAGCGCGGCGAC 1478
QY 1488 CCACTCCCGTACCAGCGCCACATGACACAGCGCTCCGCTCCGCTCCGCGGCTTTT 1547
Db 1479 CCACTCCCGTACCAGCGCCACATGACACAGCGCTCCGCTCCGCTCCGCGGCTTTT 1538
QY 1548 CTCTGCTTTTACCAGCACTGTGACCCCGACAGCGCTCCGCTCCGCTCCGCTCCGCG 1607
Db 1539 CTCTGCTTTTACCAGCACTGTGACCCCGACAGCGCTCCGCTCCGCTCCGCTCCGCG 1598
QY 1608 GCAGTACTGGGACCTTCCCTGGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667
Db 1599 GCAGTACTGGGACCTTCCCTGGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658
QY 1668 AGGCTTGGCCCTCAGTGGAGCTGCTCCACAGCGCTGGGCTGAGCTCAGAGGCGGAG 1727
Db 1659 AGGCTTGGCCCTCAGTGGAGCTGCTCCACAGCGCTGGGCTGAGCTCAGAGGCGGAG 1718
QY 1728 GCGAGAACTGAGTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1787
Db 1719 GCGAGAACTGAGTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1778
QY 1788 TTCATCAGACACCTCTGCGCCAGCTCAGCAGCTTCCATCAGCAGCAACCCAGGAC 1847
Db 1779 TTCATCAGACACCTCTGCGCCAGCTCAGCAGCTTCCATCAGCAGCAACCCAGGAC 1838
QY 1848 TTGGCTTCCCTCCTCAGCACTCAGAGCCATGCTCCCGAGCTGGGAGCACTCAACCT 1907
Db 1839 TTGGCTTCCCTCCTCAGCACTCAGAGCCATGCTCCCGAGCTGGGAGCACTCAACCT 1898
QY 1908 CCCCCCTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1967
Db 1899 CCCCCCTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1958
QY 1968 ATACCTGCCATACCAACCCAGGTATTAATCTCGCTGGTTTGTATTTATTTATTT 2027
Db 1959 ATACCTGCCATACCAACCCAGGTATTAATCTCGCTGGTTTGTATTTATTTATTT 2018
QY 2028 TTTTGTGTTGATTTTTTAAATAGAAATTTTCAATTTTAAAGCAAAAAA 2085
Db 2019 TTTTGTGTTGATTTTTTAAATAGAAATTTTCAATTTTAAAGCAATTTTAAAGCA 2076
```

```
RESULT 9
AAV64991
ID AAV64991 standard; cDNA; 2928 BP.
XX
AC AAV64991;
XX
DT 05-FEB-1999 (first entry)
XX
DE Human RAR-alpha cDNA.
XX
KW Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;
KW acute promyelocytic leukaemia; APL; t(15;17); translocation;
KW treatment; all-trans retinoic acid; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 103..1491
FT FT /*tag= a
FT FT /product= "RAR-alpha"
FT FT /note= "retinoic acid receptor-alpha"
XX
PN US5843642-A.
XX
PD 01-DEC-1998.
XX
PF 21-JUL-1993; 93US-0095728.
XX
PR 21-JUL-1993; 93US-0095728.
PR 22-MAR-1991; 91US-0673838.
PR 22-MAR-1991; 91US-0675084.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI Dmitrovsky E, Frankel S, Miller WH, Warrell RP;
XX
DR WPI: 1999-044563/04.
DR P-PSDB; AAW81984.
XX
PT Diagnosis of acute promyelocytic leukaemia - by detecting nucleic
PT acid encoding abnormal retinoic acid receptor-alpha
XX
PS Disclosure; Column 29-34; 38pp; English.
XX
CC This sequence encodes the human retinoic acid receptor alpha, RAR-alpha
CC which is used in a method for identifying a subject with acute
CC promyelocytic leukaemia (APL) resulting from a t(15;17) translocation who
CC will respond to treatment with all-trans retinoic acid. The protein can
CC also be used to identify a subject with indications of APL who will not
CC respond to treatment with all-trans retinoic acid.
XX
SQ Sequence 2928 BP; 582 A; 1011 C; 789 G; 546 T; 0 other;

Query Match 85.3%; Score 1778.8; DB 20; Length 2928;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1786; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 288 CTCATTGAGACCCAGAGCAGCTTCTGAGAGATAGTCCCGAGCCTCCCTCGCCACC 347
Db 279 CACCATTGAGACCCAGAGCAGCTTCTGAGAGATAGTCCCGAGCCTCCCTCGCCACC 338
QY 348 CCTCTACCCCGCATCTACAAGCCTTGTCTGTGTCAGGACAAGTCTCAGGCTACCA 407
Db 339 CCTCTACCCCGCATCTACAAGCCTTGTCTGTGTCAGGACAAGTCTCAGGCTACCA 398
QY 408 CTATGGGCTCAGGCGCTGTGAGGCTGCAAGGCTTCTTCCCGCGCAGCATCCAGAGAA 467
Db 399 CTATGGGCTCAGGCGCTGTGAGGCTGCAAGGCTTCTTCCCGCGCAGCATCCAGAGAA 458
QY 468 CATGTTGATACGCTGTCCCGGAGCAAGACTGCATCATCAACAGGTGACCGGACCC 527
Db 459 CATGTTGATACGCTGTCCCGGAGCAAGACTGCATCATCAACAGGTGACCGGACCC 518
```


XX (INNO-) INNOVIR LAB INC.
PA George ST, Goldberg AR, Pace U;
PI WPI: 1996-300650/30.
DR RNA construct(s) including ribozyme(s) and antisense
XX oligo:nucleotide(s) - for the inactivation of RNA associated with,
PT e.g. promyelocytic leukemia or follicular lymphoma
XX Example 1; Page 53-55; 81pp: English.
PS PML-RAR-alpha RNA (AAT33246) is associated with acute promyelocytic
CC leukaemia (APL). APL is characterised by a balanced and
CC reciprocal translocation between the long arms of chromosomes 15
CC and 17, resulting in a fusion of the retinoic acid receptor gene
CC (RAR-alpha, see also AAT33260) and a gene for a putative
CC transcription factor, PML. Methods for treating APL involve the
CC use of ribozymes, external guide sequences and antisense
CC oligonucleotides (see also AAT33245 and AAT33248-57) that specifically
CC cleave the PML-RAR-alpha fusion mRNA (see also AAT33251) but not
CC wild-type RAR-alpha mRNA.
XX Sequence 3511 BP; 713 A; 1168 C; 1038 G; 591 U; 1 other;
SQ

Query Match 85.2%; Score 1777.4; DB 17; Length 3511;
Best Local Similarity 81.5%; Pred. No. 0;
Matches 1456; Conservative 325; Mismatches 6; Indels 0; Gaps 0;

QY 290 CCATTGAGACCCAGACGACGCTTCTGAGAGATAGTCCCGAGCCCTCCCTCCGACCCC 349
DB 1725 CCAUAGAGACCCAGACGACGCTTCTGAGAGATAGTCCCGAGCCCTCCCTCCGACCCC 1784
QY 350 CTCTACCCGCGATCTACAGCCCTTGTCTGTCTGAGAGATAGTCCCTCAGGTACCACT 409
DB 1785 CUCUACCCGCGAUCACAGCCUUGCUUGUCUGAGGACGAGGACGAGGACGAGGACG 1844
QY 410 ATGGGGTGACGCGCTGTAGGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGGAACA 469
DB 1845 AUGGGGUCACGCGCUGAGGAGGCGUGCAAGGGUUCUUCGCGCAGCAUCCAGGAACA 1904
QY 470 TGGGTGTACAGTGTACCGGACAGAACTGCATCATCAACAGGTGACCGGACCCCT 529
DB 1905 UGGUGACACGUGUACCGGGACAGAACUGCAUCAACAGGAGGACCGGACCGCU 1964
QY 530 GCCAGTACTCCGAGTGCAGAGTGTCTTGAAGTGGGCATGTCCAAAGAGTCTGTGAGAA 589
DB 1965 GCCAGUACGCGACUGCAGAGUGCUUUGAGUGGCGCAUGUCCAGAGGAGUCUGAGAA 2024
QY 590 ACGACCGAAACAGAAAGAGAGGAGGTGCCAAGCCCGAGTGTCTGTGAGAGTACACGC 649
DB 2025 ACGACCGAAACAGAAAGAGAGGAGGUGGCCAAGCCCGAGUCUGAGAGCAACGC 2084
QY 650 TGACCCGAGGTGGGGAGCTCATTGAGAGGTGGCGCAAGCGCACCAAGAACTTCC 709
DB 2085 UGACCCGAGGAGGUGGAGUCUUGAGAGGUGGCGCAAGGCGCACCAAGAAACCUUC 2144
QY 710 CTGCCCTCTGCCAGTGGGCAATACACTACGAACACAGCTCAGAACACGCTGTCTCTC 769
DB 2145 CUGCCUUGCCAGCUGGCGAAUACACUACGAAACAGCUCAGAAACAGGUGUCUC 2204
QY 770 TGGACATTGACCTCTGGGACAAAGTTTCACTGTAAGTCTTCCACCAAGTGCATCAAGACTG 829
DB 2205 UGGACAUUAGACCUUGGACAAAGUUCAGUGAUCUCUCCACCAAGUGCAUCAUUAAGACUG 2264
QY 830 TGGAGTTCGCCAAGAGCTGCCGCTTCCACCCTCACCCTCAGGACGATCACCC 889
DB 2265 UGGAGUUGCCCAAGAGCUGGCGGCUUCACCAACCCUCCACCAUCCGCGGACCAUCC 2324
QY 890 TCCTCAAGGCTCCCTGCTGGACATCTCTGATCTCTGGGATCTGACGCGGTACACGCCCG 949
DB 2325 UCCUCAAGGCGCUGGACUCCUGAUCUUGGCGGAGUUGCAGCGCGGACGCGGCCCG 2384

QY 950 AGCAGGACCATGACCTTCTCGAGCGGCTGACCTGAACCGGACCCAGATGCAACAG 1009
DB 2385 AGCAGGACCATGACCTTCTCGAGCGGCTGACCTGAACCGGACCCAGATGCAACAG 2444
QY 1010 CTGGCTTCGGCCCCCTCACCGACCTGCTTTCCTTCGCGCAACAGCTGTGCCCCCTGG 1069
DB 2445 CUGGCUUGGCCCCCTCACCGACCTGCTTTCCTTCGCGCAACAGCTGTGCCCCCTGG 2504
QY 1070 AGATGGATGATGGGAGACGCGGCTGTACGCGCATCTGCTCTCATCTGCGGAGACCGCC 1129
DB 2505 AGAUGGAUGAUGCGGAGACGCGGCTGTACGCGCATCTGCTCTCATCTGCGGAGACCGCC 2564
QY 1130 AGGACCTGGAGACGCGGACCGGCTGACATGCTGCAGGACCGCTGCTGAGGCGCTAA 1189
DB 2565 AGGACCTGGAGACGCGGACCGGCTGACATGCTGCAGGACCGCTGCTGAGGCGCTAA 2624
QY 1190 AGGTCTACGTGCGGAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1249
DB 2625 AGGUCUACGUGGCGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2684
QY 1250 AGATTACTGACCTGCGGAAGCATCAGCGCCCAAGGGGGGTGAGCGGCTGATCATCGTGAAGA 1309
DB 2685 AGAUUACUGACCGCGGAAGCAUCAGCGCCCAAGGGGGCGGCGGCGGCGGCGGCGGCGG 2744
QY 1310 TGGAGATCCCGGGTCCATGCGCGCTCTCATCAGGAAATTTGGAGAACTCAGAGGCC 1369
DB 2745 UGGAGAUCCCGGGTCCATGCGCGCTCTCATCAGGAAATTTGGAGAACTCAGAGGCC 2804
QY 1370 TGGACACTGTAGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1429
DB 2805 UGGACACUCUGAGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2864
QY 1430 CGCCAGGACGCTGTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1489
DB 2865 CGCCAGGACGCTGTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2924
QY 1490 ACTCCCGGTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1549
DB 2925 ACUCCCGUGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2984
QY 1550 CTCGCTTTCACGACCATGTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1609
DB 2985 CUGCCUUCUACCGACCAUGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3044
QY 1610 AGTACTGGGACCTTCCCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1669
DB 3045 AGUACUGGGACCUUCCUUGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3104
QY 1670 GCCTGGGCGCTCAGTGGACCTGCTGCTCCACAGCTGGGCTGAGTCTGAGAGCGGAGGC 1729
DB 3105 GCGUGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3164
QY 1730 CAGGAAGTGTAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1789
DB 3165 CAGGAAGTGTAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3224
QY 1790 CATCAAGACACCGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1849
DB 3225 CAUCAAGACACCGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3284
QY 1850 GGCTCCCGCATCTCAGAACTCAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1909
DB 3285 GGCUCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3344
QY 1910 CCCTGCTCGGTTGGTGTGACAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1969
DB 3345 CCCUGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3404
QY 1970 ACCTGCGCATACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2029
DB 3405 ACCUGCGCATACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3464

Qy	2030	TTGTTTGATTTTTTATAAGAATTTCATTTAAGCACAAAAA	2076
	:	: :	
Dd	3465	UGUUUUGAUUUUUUAUAAGAAUUUCACUUUUAAGCAUUAAAAA	3511

RESULT 11
AAQ29338
ID AAQ29338 standard; cDNA; 2928 BP.

AA
AC
XX
DT
XX
DE

AAQ29338;
09-MAR-1993 (first entry)
RAR-alpha gene.

KW	Retinoic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;
KW	APL; translocation; chromosome 17; chromosome 15; PCR; primer;
KW	t(15;17)(q21;q11-22); breakpoint; polymerase chain reaction; ss.
XX	
OS	Synthetic.

	Key	Location/Qualifiers
FH		
FT	CDS	103..1488
FT		/*taq= a

PN WO9216660-A.

01-OCT-1992.

23-MAR-1992; 92WO-US02320.

PR 22-MAR-1991; 91US-0673838.

PR 22-MAR-1991; 91US-0675084.

PA (SLOK) SLOAN KETTERING INST CANCER.

PI Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;
PI Warrell RP;

DR WPI; 1992-349240/42.

DR P-PSDB; AAR27534.

PT Marker for acute promyelocytic leukaemia and other neoplasias -
PT comprising nucleic acid and encoded abnormal retinoic acid
PT receptor-alpha receptor

PS Disclosure; Page 43-46; 84pp; English.

The sequence given represents the nucleic acid sequence of the retinoic acid receptor (RAR)-alpha gene. This gene is disrupted in a translocation of a portion of the long arm of chromosome 17 onto the long arm of chromosome 15 (t(15;17)(q21;q11-22)). This causes a fusion between RAR-alpha and myl which is characteristic of acute promyelocytic leukemia (APL). The breakpoint region has been cloned and it has been shown that DNA rearrangements are clustered in the region of the first intron of RAR-alpha. This sequence was isolated by polymerase chain reaction (PCR). The primers used for amplification of this sequence can also be used to amplify the translocated region.

Sequence 2928 BP; 582 A; 1012 C; 789 G; 545 T; 0 other;

Query Match	85.2%	Score 1777.2;	DB 13;	Length 2928;
Best Local Similarity	99.3%;	Pred. No. 0;		
Matches 1785;	Conservative	0;	Mismatches	13;
			Indels	0;
			Gaps	0;

QY 288 CTCCATTGAGACCCAGAGCAGCAGTCTGAAGAGATAGTGCCCAGCCCTCCCTCGCCACC 347

Db 279 CACCATTGAGACCCAGAGCAGCAGTTCTGAAGAGATAGTCCCAGCCCTCCCTCGCCACC 338

348 CCCTCTACCCCGCATCTACAAGCCTTGCTTTGTCTGTCTCAGGACAAGTCCTCAGGGTACCA 407

Db 339 CCCTCTACCCCGCATCTACAAGCCTTGCTTTGTCTGTCTCAGGACAAGTCTCAGGCTACCA 398

Db 1059 GGAGTGGATGATCGGAGACGGGGCTGCTCAGCGCATCTGCTCATCTCGGAGACCG 1118
Qy 1128 CCAGGACCTGGAGACCGGACCGGGTGCATGCTGCAGGACCGCTGCTGGAGCGCT 1187
Db 1119 CCAGGACCTGGAGACCGGACCGGGTGCATGCTGCAGGACCGCTGCTGGAGCGCT 1178
Qy 1188 AAAGTGTACGTGCGGAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAT 1247
Db 1179 AAAGTGTACGTGCGGAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAT 1238
Qy 1248 GAAGTATTACTGACCTCGAAGCATCAGCGCCAGGCGGTGAGCGGTGATCAGCTGAA 1307
Db 1239 GAAGTATTACTGACCTCGAAGCATCAGCGCCAGGCGGTGAGCGGTGATCAGCTGAA 1298
Qy 1308 GATGAGATCCCGGGTCCATCGCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGG 1367
Db 1299 GATGAGATCCCGGGTCCATCGCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGG 1358
Qy 1368 CTTGACACTCTGAGCGGACCGGGGGTGGGGGGCGGAGCGGGTGGCTGCCGCC 1427
Db 1359 CTTGACACTCTGAGCGGACCGGGGGTGGGGGGCGGAGCGGGTGGCTGCCGCC 1418
Qy 1428 CCGCCAGGACCTGTAGCCCCAGCTCAGCCCCAGCTCCACAGAGAGCAGCCCGCCAC 1487
Db 1419 CCGCCAGGACCTGTAGCCCCAGCTCAGCCCCAGCTCCACAGAGAGCAGCCCGCCAC 1478
Qy 1488 CCACCTCCCGGTGACCGCCACCGCCACATGACACAGCCCTCGCCCTCCCGCGGCTTT 1547
Db 1479 CCACCTCCCGGTGACCGCCACCGCCACATGACACAGCCCTCGCCCTCCCGCGGCTTT 1538
Qy 1548 CTCCTGCTTTTACCGACCATGTGACCCCGCAGCCCTGCCCCACCTGCCCTCCCGG 1607
Db 1539 CTCCTGCTTTTACCGACCATGTGACCCCGCAGCCCTGCCCCACCTGCCCTCCCGG 1598
Qy 1608 GCAGTACTGGGACCTTCCTCGGGGACGGGGAGGAGGAGCAGCAGCTCTTGACAG 1667
Db 1599 GCAGTACTGGGACCTTCCTCGGGGACGGGGAGGAGGAGGAGCAGCAGCTCTTGACAG 1658
Qy 1668 AGGCTTGGGCTTCACTGAGTGGCTGCTGCTCCACAGCCTGGGCTGACGTCAGAGGCGGAG 1727
Db 1659 AGGCTTGGGCTTCACTGAGTGGCTGCTGCTCCACAGCCTGGGCTGACGTCAGAGGCGGAG 1718
Qy 1728 GCCAGAACTAGTGGAGCCCTGCTGCTGGGTCTCAGGATGGGTCTCGGGGCTCTGCTG 1787
Db 1719 GCCAGAACTAGTGGAGCCCTGCTGCTGGGTCTCAGGATGGGTCTCGGGGCTCTGCTG 1778
Qy 1788 TTCATCAGACACCCCTCTGCCCCAGCTCAGCAGCTTTCATCAGCAGAAACGCCAGGAC 1847
Db 1779 TTCATCAGACACCCCTCTGCCCCAGCTCAGCAGCTTTCATCAGCAGAAACGCCAGGAC 1838
Qy 1848 TTGGCTCCCGCATCTCAGAACTCAGAGCCATGCTCCCGAGCTGGGAACTCAACCT 1907
Db 1839 TTGGCTCCCGCATCTCAGAACTCAGAGCCATGCTCCCGAGCTGGGAACTCAACCT 1898
Qy 1908 CCCCCCTGCTCGGTGTGACAGAGGGGTGGGAGAGGGGGGGGGTTCGCCCTGTAC 1967
Db 1899 CCCCCCTGCTCGGTGTGACAGAGGGGTGGGAGAGGGGGGGGGTTCGCCCTGTAC 1958
Qy 1968 ATACCTGCCATACCAACCCAGGTATTAATTTCTCGCTGTTTTTTTATTTAATTT 2027
Db 1959 ATACCTGCCATACCAACCCAGGTATTAATTTCTCGCTGTTTTTTTATTTAATTT 2018
Qy 2028 TTTTGTGTTTTTTTATTAAGATTTTTCATTTTAAAGCAAAAAA 2085
Db 2019 TTTTGTGTTTTTTTATTAAGATTTTTCATTTTAAAGCAATTTATCTAAGGAA 2076

RESULT 13
AAQ81476
ID AAQ81476 standard; cDNA to mRNA; 2940 BP.
XX
AC AAQ81476;

XX 01-SEP-1995 (first entry)
DT RAR-alpha clone lambda-KIR.
XX RAR-alpha; retinoic acid receptor alpha; hematopoietic; stem cell;
XX differentiation; ds.
KW Homo sapiens.
OS
XX Key Location/Qualifiers
FH 103..1491
FT /*tag- a
XX
PN W09504143-A.
XX
PD 09-FEB-1995.
XX
XX 28-JUL-1994; 94WO-US08450.
XX
XX 28-JUL-1993; 93US-0099242.
PR (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
XX Collins SJ, Tsai S;
PI
XX WPI; 1995-082227/11.
DR P-PSDB; AAR68023.
XX
PT New haematopoietic stem cell lines with specific differentiation
PT properties - made by transfected stem cells with nucleic acid
PT encoding dominant negative suppressor of the retinoic acid
PT receptor alpha, useful e.g. for haematopoietic reconstitution
XX
XX Disclosure; Page 55-56; 100pp; English.
XX
XX An RAR-alpha cDNA (given in AAQ81477) was prepared that contained a
CC truncation of sequences coding for the C-terminal 59 amino acids
CC and part the 3' UTR of the wild-type human sequence (AAQ81476).
CC This truncated cDNA, designated RAR-alpha-403, encodes a protein
CC (AAR68024) containing the N-terminus, DNA-binding domain and part of
CC the hormone-binding domain of RAR-alpha (AAR68023).
XX
SQ Sequence 2940 BP; 594 A; 1012 C; 788 G; 546 T; 0 other;
Query Match 85.2%; Score 1777.2; DB 16; Length 2940;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1785; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 288 CTCATTGAGACCCAGAGCAGCAGTTCCTGAAGAGATAGTGCACCGCTCCCTCGCCACC 347
Db 279 CACCTTGAAGACCCAGAGCAGCAGTTCCTGAAGAGATAGTGCACCGCTCCCTCGCCACC 338
Qy 348 CCCTTACCCCGCATCTACAAGCCTTGTGTTGTCTGTTCAGACAAAGTCTCTCAGGCTACCA 407
Db* 339 CCCTTACCCCGCATCTACAAGCCTTGTGTTGTCTGTTCAGACAAAGTCTCTCAGGCTACCA 398
Qy 408 CTATGGGCTCAGCCCTGTGAGGGCTGCAAGGGCTTCTTCGGCGCAGCATCCAGAAGAA 467
Db 399 CTATGGGCTCAGCCCTGTGAGGGCTGCAAGGGCTTCTTCGGCGCAGCATCCAGAAGAA 458
Qy 468 CATGGGTACACCTGTTCACCGGACAAAGAACTGCATCATCAACAAGTGTACCCGAAACCC 527
Db 459 CATGGGTACACCTGTTCACCGGACAAAGAACTGCATCATCAACAAGTGTACCCGAAACCC 518
Qy 528 CTGCCAGTACTGCCGACTGCAGAAAGTGTGTTGAAGTGGGATGTCTCAAGGAGTCTGTGAG 587
Db 519 CTGCCAGTACTGCCGACTGCAGAAAGTGTGTTGAAGTGGGATGTCTCAAGGAGTCTGTGAG 578
Qy 588 AAACGACCGAAACAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTGTGAGAGCTACAC 647
Db 579 AAACGACCGAAACAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTGTGAGAGCTACAC 638

disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 3052 BP; 613 A; 1020 C; 894 G; 525 T; 0 other;

Query Match 81.9%; Score 1707.4; DB 23; Length 3052;
Best Local Similarity 99.2%; Pred. No. 2.2e-309;
Matches 1789; Conservative 0; Mismatches 6; Indels 8; Gaps 7;

QY 290 CCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTCCCGACGCTCCCTCGCCACCCC 349
|||||
Db 1250 CCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTCCCGACGCTCCCTCGCCACCCC 1309
|||||

QY 350 CTCFACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAGTCTCAGGCTACCACT 409
|||||
Db 1310 CTCFACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAGTCTCAGGCTACCACT 1369
|||||

QY 410 ATGGGGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTCCGCGCAGCATCAAGAGTACAGAGAAC 469
|||||
Db 1370 ATGGGGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTCCGCGCAGCATCAAGAGTACAGAGAAC 1429
|||||

QY 470 TGGTGTACACCTGTCACCGGACAGAACTGCATCATCAAGAGTACAGAGTACAGAGTACAG 529
|||||
Db 1430 TGGTGTACACCTGTCACCGGACAGAACTGCATCATCAAGAGTACAGAGTACAGAGTACAG 1489
|||||

QY 530 GCCAGTACTGCGGACTGCAGAGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAGAA 589
|||||
Db 1490 GCCAGTACTGCGGACTGCAGAGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAGAA 1549
|||||

QY 590 ACGACCAAGCAAGAGAGAGTGGCCAAAGCGGAGTGTCTGAGAGTACAGC 649
|||||
Db 1550 ACGACCAAGCAAGAGAGAGTGGCCAAAGCGGAGTGTCTGAGAGTACAGC 1609
|||||

QY 650 TGACCGCGAGGTGGGGAGCTATTGAGAGTGGCCAAAGCGGAGTGTCTGAGAGTACAGC 709
|||||
Db 1610 TGACCGCGAGGTGGGGAGCTATTGAGAGTGGCCAAAGCGGAGTGTCTGAGAGTACAGC 1669
|||||

QY 710 CTGCGCTCTGCGAGTGGGCAATACACTAGCAACAGCTCAGCAACAGTGTCTCTC 769
|||||
Db 1670 CTGCGCTCTGCGAGTGGGCAATACACTAGCAACAGCTCAGCAACAGTGTCTCTC 1729
|||||

QY 770 TGGACATTTGACCTTGGGACAGTTTCACTTCAACAGTGTCTTCAAGAGT 829
|||||
Db 1730 TGGACATTTGACCTTGGGACAGTTTCACTTCAACAGTGTCTTCAAGAGT 1789
|||||

QY 830 TGAGATTGCGCAAGCAGCTCCCGGCTTCAACCCCTCACCATCGCGACAGATCACCC 889
|||||
Db 1790 TGAGATTGCGCAAGCAGCTCCCGGCTTCAACCCCTCACCATCGCGACAGATCACCC 1849
|||||

QY 890 TCCTCAAGGTGCTGCTGGACATCTGTATCTGCGGATCTGACGCGGTACAGCGCG 949
|||||
Db 1850 TCCTCAAGGTGCTGCTGGACATCTGTATCTGCGGATCTGACGCGGTACAGCGCG 1909
|||||

QY 950 AGCAGGACACCATGACCTTCTCGGACGGGTGACCCCTGAAACCGGACCGATGACAAAG 1009
|||||
Db 1910 AGCAGGACACCATGACCTTCTCGGACGGGTGACCCCTGAAACCGGACCGATGACAAAG 1969
|||||

QY 1010 CTGGCTTGGCGCCCTCACCAGCTGCTTGGCTTGGCCCAACAGCTGCTGCCCTGG 1069
|||||
Db 1970 CTGGCTTGGCGCCCTCACCAGCTGCTTGGCTTGGCCCAACAGCTGCTGCCCTGG 2029
|||||

QY 1070 AGATGGATGATGGGAGACGGGCTGCTCAGCGCATCTGCCTCATCTGGGAGACCGCC 1129
|||||
Db 2030 AGATGGATGATGGGAGACGGGCTGCTCAGCGCATCTGCCTCATCTGGGAGACCGCC 2089
|||||

QY 1130 AGGACCTGAGAGACCGCGGACATGCTGCAGGAGCGCTCTCTGAGAGCGGTAA 1189
|||||

Db 2090 AGGACCTGAGAGACCGCGGACATGCTGCAGGAGCGCTCTCTGAGAGCGGTAA 2149
|||||

QY 1190 AGGCTTACCTGCGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGTAATGA 1249
|||||

Db 2150 AGGCTTACCTGCGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGTAATGA 2209
|||||

QY 1250 AGATTACTACCTGCGGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGATCAGCTGAAGA 1309
|||||

Db 2210 AGATTACTACCTGCGGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGATCAGCTGAAGA 2269
|||||

QY 1310 TGGAGATCCCGGCTCCATGCGCGCTTCTCATCAGGAAATGTTGGAGAACTCAGAGGGCC 1369
|||||

Db 2270 TGGAGATCCCGGCTCCATGCGCGCTTCTCATCAGGAAATGTTGGAGAACTCAGAGGGCC 2329
|||||

QY 1370 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGGGAGCGGGGTGGCTGCCCGCCC 1429
|||||

Db 2330 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGGGAGCGGGGTGGCTGCCCGCCC 2389
|||||

QY 1430 CGCCAGGACAGCTGTAGCCCGCAGCTCAGCGCCAGCTTCCAAAGAGCAGCGCCGCGCACCC 1489
|||||

Db 2390 CGCCAGGACAGCTGTAGCCCGCAGCTCAGCGCCAGCTTCCAAAGAGCAGCGCCGCGCACCC 2449
|||||

QY 1490 ACTCCCGTGTAGCGCCCGCAGCATGACAGCGCTTCCCGCTCGCGCCCGCGCTTCTTC 1548
|||||

Db 2450 ACTCCCGTGTAGCGCCCGCAGCATGACAGCGCTTCCCGCTCGCGCCCGCGCTTCTTC 2509
|||||

QY 1549 TCTGCTTCTTACCGACCATGTGACCCC--GCACAGCCCTGCCCGCCCGCCT--GCCCTCCC 1605
|||||

Db 2510 TCTGCTTCTTACCGACCATGTGACCCC--GCACAGCCCTGCCCGCCCGCCT--GCCCTCCC 2569
|||||

QY 1606 GGGCAGTACTGGGGACCTTCCCTGGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1665
|||||

Db 2570 GGGCAGTACTGGGGACCTTCCCTGGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2629
|||||

QY 1666 AGAGGCTTGGGCTTCTGAGTGGCTGCTCTCCAGCGCTTGGGCTGAGCTCAGAGGCGG 1725
|||||

Db 2630 AGAGGCTTGGGCTTCTGAGTGGCTGCTCTCCAGCGCTTGGGCTGAGCTCAGAGGCGG 2689
|||||

QY 1726 AGG--CCAGGAACTGAGTGGGCGGCTGGT--CCTGGGTCTCAGAGTGGGT--CCTGGGGGCG 1782
|||||

Db 2690 AGGCGCAGGAACTGAGTGGGCGGCTGGTCCCTGGGTCTCAGGATGGGTCTGGGGGCG 2749
|||||

QY 1783 TCGTCTTCTATCAAGACACCCCTCTGCCAGCTCACCACATCTCATCACCAGCAACGCG 1842
|||||

Db 2750 TCGTCTTCTATCAAGACACCCCTCTGCCAGCTCACCACATCTCATCACCAGCAACGCG 2809
|||||

QY 1843 AGGACTTGGCTTCCCGCATCTCAGAACTCACAAGCCATT--GCTTCCCGCAGCTGGGGAACT 1901
|||||

Db 2810 AGGACTTGGCTTCCCGCATCTCAGAACTCACAAGCCATTGGCTTCCCGCAGCTGGGGAACT 2869
|||||

QY 1902 CAACCTCCCGCTGCTCGGTGGTGGAGAGGGGTTGGAGAGGGGCGGGGGTTCGCC 1961
|||||

Db 2870 CAACCTCCCGCTGCTCGGTGGTGGAGAGGGGTTGGAGAGGGGCGGGGGTTCGCC 2929
|||||

QY 1962 CTGTACATACCTGCTATACCAACCCAGGTATTATCTCGCTGCTTGTGTTTATTT 2021
|||||

Db 2930 CTGTACATACCTGCTATACCAACCCAGGTATTATCTCGCTGCTTGTGTTTATTT 2989
|||||

QY 2022 TAATTTTTTGTGTTTTTAAAGAAATTTTCAATTTTAAAGCAAAAAA 2081
|||||

Db 2990 TAATTTTTTGTGTTTTTAAAGAAATTTTCAATTTTAAAGCAAAAAA 3049
|||||

QY 2082 AAA 2084
|||

Db 3050 AAA 3052

RESULT 15
AAS83048
ID AAS83048 standard; cdna; 2930 BP.
XX

AAS83048;
 13-FEB-2002 (first entry)
 DNA encoding novel human diagnostic protein #18852.
 Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS
 Homo sapiens.
 WO200175067-A2.
 11-OCT-2001.
 30-MAR-2001; 2001WO-US08631.
 31-MAR-2000; 2000US-0540217.
 23-AUG-2000; 2000US-0649167.
 (HYSB-) HYSEQ INC.
 Drmanac RT, Liu C, Tang YT;
 WPI; 2001-639362/73.
 P-PSDB; ABG18861.
 New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity -
 Claim 1; SEQ ID NO 18852; 103pp; English.
 The invention relates to isolated polynucleotide (I) and
 polypeptide (II) sequences. (I) is useful as hybridisation probes,
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 and gene mapping, and in recombinant production of (II). The
 polynucleotides are also used in diagnostics as expressed sequence tags
 for identifying expressed genes. (I) is useful in gene therapy techniques
 to restore normal activity of (II) or to treat disease states involving
 (II). (II) is useful for generating antibodies against it, detecting or
 quantitating a polypeptide in tissue, as molecular weight markers and as
 a food supplement. (II) and its binding partners are useful in medical
 imaging of sites expressing (II). (I) and (II) are useful for treating
 disorders involving aberrant protein expression or biological activity.
 The polypeptide and polynucleotide sequences have applications in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits to assess biodiversity
 and to produce other types of data and products dependent on DNA and
 amino acid sequences. AAS64197-AAS94564 represent novel human
 diagnostic coding sequences of the invention.
 Note: the sequence data for this patent did not appear in the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.
 Sequence 2930 BP; 565 A; 1008 C; 807 G; 550 T; 0 other;

Query Match 73.8%; Score 1540.2; DB 23; Length 2930;
 Best Local Similarity 97.2%; Pred. No. 3.5e-278;
 Matches 1770; Conservative 0; Mismatches 28; Indels 23; Gaps 19;

QY	468	CATGTTGTACAGTGTCTACCGGCAAGAAGTGCATCATCAACAAGGTGACCCGGACCC	527
DB	459	CATGTTGTACAGTGTCTACCGGCAAGAAGTGCATCATCAACAAGGTGACCCGGACCC	518
QY	528	CTGCCAGTACTGCCGACTGCAGAGTGTCTTTGAAGTGGGCATGTCCAAAGAGTCTCTGAG	587
DB	519	CTGCCAGTACTGCCGACTGCAGAGTGTCTTTGAAGTGGGCATGTCCAAAGAGTCTCTGAG	578
QY	588	AAACGACCGAAACAAAGAAGAAGAGAGGTGCCCAAGCCCGAGTCTCTGAGAGTACAC	647
DB	579	AAACGACCGAAACAAAGAAGAAGAGAGGTGCCCAAGCCCGAGTCTCTGAGAGTACAC	638
QY	648	GCTGACCGCGGAGGTGGGGAGTCTATTGAGAAGTGGCGCAAGCGCACCA-GGAACCT	706
DB	639	GCTGACCGCGGAGGTGGGGAGTCTATTGAGAAGTGGCGCAAGCGCACCAAGGAACT	698
QY	707	TCCCTGCCCTCTGCCAGCTGGGCAAAATACACTAC-GAACAACAGC-TCAAGAACAAGTG-	763
DB	699	TCCCTGCCCTCTGCCAGCTGGGCAAAATACACTACAGGAACAACAGCTTCAGAACAAGTGT	758
QY	764	TCTCTCTGGACATTGACCTCTGGGACAAGTTCAAGTGAAGTCTCCACCAAGTGCATTA	823
DB	759	TCTCTCTGGACATTGACCTCTGGGACAAGTTCAAGTGAAGTCTCCACCAAGTGCATTA	818
QY	824	AGACTGTGG-AGTTGCCCAAG--CAGCTGCCGGCTTCAACACCTCACCATC-GCCGAC	879
DB	819	AAAATGGGAGGTCGCCAAGGCGAGGCTGCCGGCTTCAACACCTCACCATCAGCCGAC	878
QY	880	CAGATACCCCTCCCAA-GGCTGCTGCCT-GGACATCCTGATCTCTGGGATCTGCACGC	937
DB	879	CAGATACCCCTCCCAAAGGCTGCTGCTGGGATCCTGATCTCTGGGATCTGCACGC	938
QY	938	GGTACACGCCGAG--CAGGACACCATGACCTTC-TCGGAGCGGCTGACCTGAAACC-GG	993
DB	939	GGTACACGCCGAGGCGGACCATGACCTTCTTGGAGCGGCTGACCTGAAACCGG	998
QY	994	ACCAGATGACACAGCTGG--CTTGGCCCCCTCACGACCTGTCTTTGCC-TTGGCC	1050
DB	999	ACCAGATGACACAGCTGGGCTTCCGCCCTTTCACGACCTGTCTTTGCCCTTCCGC	1058
QY	1051	AACAGCTGTCTGCCCTT-GGAGATGGATGATG--CGGAGCGGCTGCTCAGCGCCATC	1107
DB	1059	AACAGCTGTCTGCCCTTGGGAGATGGATGAGGCGGAGACGGGCTGCTCAGCGCCATC	1118
QY	1108	TGCTCATCTCGGAGACCGCCAGGACCTGGAGCAGCGGACCGGCTGAGATGCTGCAG	1167
DB	1119	TGCTCATCTCGGAGACCGCCAGGACCTGGAGCAGCGGACCGGCTGAGATGCTGCAG	1178
QY	1168	GAGCGCTGTGGAGCGCTAAAGTCTACGTGCGGAAGCGGAGCGCCAGCCGCCAC	1227
DB	1179	GAGCGCTGTGGAGCGCTAAAGTCTACGTGCGGAAGCGGAGCGCCAGCCGCCAC	1238
QY	1228	ATGTTCCCCAAGTGTCTAATCAAGTACTGACCTGCGGAAGCATCAGGCCCAAGGGGCT	1287
DB	1239	ATGTTCCCCAAGTGTCTAATCAAGTACTGACCTGCGGAAGCATCAGGCCCAAGGGGCT	1298
QY	1288	GAGCGGTGATCAGCTGAAGATGGAGATCCCGGCTTCCATGCGGCTTCTCATCCAGAA	1347
DB	1299	GAGCGGTGATCAGCTGAAGATGGAGATCCCGGCTTCCATGCGGCTTCTCATCCAGAA	1358
QY	1348	ATTTTGAGAACTCAGAGGCTTGGACACTCTTGAGCGGACAGCGGGGGTGGGGCGG	1407
DB	1359	ATTTTGAGAACTCAGAGGCTTGGACACTCTTGAGCGGACAGCGGGGGTGGGGCGG	1418
QY	1408	GAGGGGGTGGCTTCCCCCCCCGCCA-GGAGCTGTAGCCCCAGCTCAGCCCCAGCTC	1466
DB	1419	GAGGGGGTGGCTTGGCCCCCCCCGCCAGGAGCTGTAGCCCCAGCTCAGCCCCAGCTC	1478
QY	1467	CAACAGAAGCAGCCCGGCCACCCACTCCCGTGTACCGCCGACGACCATGACACAGCC	1526
DB	1479	CAACAGAAGCAGCCCGGCCACCCACTCCCGTGTACCGCCGACGACCATGACACAGCC	1538

QY 1527 TCGCCCTCCGCCCGCGGCTTTCTCTGCTTTCTACCGACCATGTGACCCCGCACAGCCC 1586
|||||
Db 1539 TCGCCCTCCGCCCGCGGCTTTCTCTGCTTTCTACCGACCATGTGACCCCGCACAGCCC 1598
|||||
QY 1587 TCGCCCCACCTGCGCCT-CCGCGCAGTACTGGGGACCTTCCCTGGGGGACGGGAGGGAG 1645
|||||
Db 1599 TCGCCCCACCTGCGCCTGCGCGGCAGTACTGGGGACCTTCCCTGGGGGACGGGAGGGAG 1658
|||||
QY 1646 GAGGCAGGACTCCTTGGACAGAGGCTGGGCCCTCAGTGGAGCTGCCTGCCACAGCC 1705
|||||
Db 1659 GAGGCAGGACTCCTTGGACAGAGGCTGGGCCCTCAGTGGAGCTGCCTGCCACAGCC 1718
|||||
QY 1706 TGGGCTGAGCTCAGAGGCCGAGCCAGGAACCTGAGTGAGGCCCTGGTCTGGGTCTCAG 1765
|||||
Db 1719 TGGGCTGAGCTCAGAGGCCGAGCCAGGAACCTGAGTGAGGCCCTGGTCTGGGTCTCAG 1778
|||||
QY 1766 GATGGTCTCGGGGCTCGTGTTCATCAGACACCCCTCTGCCAGCTCACCACATCTT 1825
|||||
Db 1779 GATGGTCTCGGGGCTCGTGTTCATCAGACACCCCTCTGCCAGCTCACCACATCTT 1838
|||||
QY 1826 CATCACGACCAAGCCAGGACTTGGTCCGCCATCCTC-AGAACTCACAAGCCATTGCT 1884
|||||
Db 1839 CATCACGACCAAGCCAGGACTTGGTCCGCCATCCTCAGAACTCACAAGCCATTGCT 1898
|||||
QY 1885 CCCCAGCTGGGGAACCTCAACCTCCGCCCTGCGCTGGTGGTGACAGAGGGGTGGGACA 1944
|||||
Db 1899 CCCCAGCTGGGGAACCTCAACCTCCGCCCTGCGCTGGTGGTGACAGAGGGGTGGGACA 1958
|||||
QY 1945 GGGCGGGGGGTTCGCCCTGTACATACCCCTGCCATACCAACCCAGGTATTATTCGCG 2004
|||||
Db 1959 GGGCGGGGGGTTCGCCCTGTACATACCCCTGCCATACCAACCCAGGTATTATTCGCG 2018
|||||
QY 2005 TGGTTTGTATTTATTTATTTTGTGTTTGTATTTTATTAAGAATTTTCATTTTA 2064
|||||
Db 2019 TGGTTTGTATTTATTTATTTTGTGTTTGTATTTTATTAAGAATTTTCATTTTA 2078
|||||
QY 2065 AGCACAAAAAATAAAAAA 2085
|||||
Db 2079 AGCACATTTATCTGAAGGAA 2099
|||||

Search completed: March 29, 2003, 13:27:12
Job time : 447 secs

pred. No. is the number of results predicted by chance to have a

RESULT 1	HSMB05408	2300 bp	mRNA	linear	PRI 12-JUL-2002
LOCUS	HSMB05408				
DEFINITION	Homo sapiens mRNA; cdna DKEZp761C0417 (from clone DKEZp761C0417).				
ACCESSION	AL834159				
VERSION	AL834159.1	GI:21739638			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2300)				
AUTHORS	Ansorge, W., Winkner, U., Mewes, H.W., Weil, B. and Wiemann, S.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY				

pred. No. is the number of results predicted by chance to have a

COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DFPZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp761C0417) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES

source

1. 2300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp761C0417"
/tissue_type="amygdala"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"

polyA_site 462 a 765 c 656 g 417 t
BASE COUNT
ORIGIN

Query Match 95.6%; Score 1995; DB 9; Length 2300;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2058; Conservative 0; Mismatches 10; Indels 7; Gaps 5;

QY	11	GGGCGAGGGGACGCTCTCTCCGCCAGCTGCTGCTCGGATGGCGCGCGCGC--TGA	68
DB	228	GGGCGAGGGGACGCTCTCTCTCCGCCAGCTGCTGCTCGGATGGCGCGCGCGCTGAG	287
QY	69	GTGACGGGGGGCGGCGCGAGACTTCCAGCTCGGACCTCTT--GCCTTCGAGGGGAAAG	126
DB	288	TTGACGGGGGGCGGCGCGAGACTTCCAGCTCGGACCTCTTGGCTTCGAGGGGAAAG	347
QY	127	ATGTACGAGAGTGTAGAGTGGGGGTCCACCCCTAATCCCTTCCTAGTGGTGGATTT	186
DB	348	ATGTACGAGAGTGTAGAGTGGGGGTCCACCCCTAATCCCTTCCTAGTGGTGGATTT	407
QY	187	TATTAACAGAACCGGGCTCTTTGCTCCAGAGAGGGGTCCCGCCCGCGGGTCCCTAC	246
DB	408	TATTAACAGAACCGGGCTCTTTGCTCCAGAGAGGGGTCCCGCCCGGGTCCCTAC	467
QY	247	TCCACCCCGCTCGGACTCGCGCTTTGGAATGGCTCAACACCTCCATTGAGACCCAGC	306
DB	468	TCCACCCCGCTCGGACTCGCGCTTTGGAATGGCTCAACACCTCCATTGAGACCCAGC	527
QY	307	AGCAGTCTGAAGAGATAGTGGCCAGCCCTCCCTCGCCACCCCTCTACCCCGCATCTAC	366
DB	528	AGCAGTCTGAAGAGATAGTGGCCAGCCCTCCCTCGCCACCCCTCTACCCCGCATCTAC	587
QY	367	AAGCCTTGCTTCTGCTCAGGACAAGTCTCAGGCTACCACTATGGGTCAGCGGCTGT	426
DB	588	AAGCCTTGCTTCTGCTCAGGACAAGTCTCAGGCTACCACTATGGGTCAGCGGCTGT	647
QY	427	GAGGGTGAAGGGCTCTTCCCGCAGCATCCAGAAGAACATGGTGTACAGTGTTCAC	486
DB	648	GAGGGTGAAGGGCTCTTCCCGCAGCATCCAGAAGAACATGGTGTACAGTGTTCAC	707
QY	487	CGGGACAAGAACTGCAATCAACAAGGTGACCCGGAACCCCTGCCAGTACTGCCAGCTG	546
DB	708	CGGGACAAGAACTGCAATCAACAAGGTGACCCGGAACCCCTGCCAGTACTGCCAGCTG	767
QY	547	CAGAAGTGTCTTCAGTGGGCTGTCCAGGAGTGTGAGAAACGACGAAACAAGAAG	606
DB	768	CAGAAGTGTCTTCAGTGGGCTGTCCAGGAGTGTGAGAAACGACGAAACAAGAAG	827
QY	607	AAGAAGGAGGTGCCAAGCCCGAGTCTCTGAGAGTACACGCTGACGCGGAGGTGGGG	666
DB	828	AAGAAGGAGGTGCCAAGCCCGAGTCTCTGAGAGTACACGCTGACGCGGAGGTGGGG	887
QY	667	GAGCTATTGAGAGGTGGCAAGCGCACACGAAACCTTCCTCCCTCTGCCAGCTG	726

DB	888	GAGCTCATTTAGAA--GTGGCAAAAGCGCACAGGAAACCTTCCTGCTCTGCCAGCTG	946
QY	727	GGCAATACACTACGAACAACAGCTCAGACACAGTGTCTCTCTGACATTGACCTCTGG	786
DB	947	GGCAATACACTACGAACAACAGCTCAGACACAGTGTCTCTCTGACATTGACCTCTGG	1006
QY	787	GACAAAGTTAGTGAATCTTCC--ACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCA	845
DB	1007	GACAAAGTTAGTGAATCTTCCAAACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCA	1066
QY	846	GCTGCCCGGCTTACCAACCTCACCATCGCCGACAGATACCCCTCTCAAGGCTGCCTG	905
DB	1067	GCTGCCCGGCTTACCAACCTCACCATCGCCGACAGATACCCCTCTCAAGGCTGCCTG	1126
QY	906	CCTGACATCTCTGATCTCGGATCTGCACGGGTACACGCGCGGAGACACCATGAC	965
DB	1127	CCTGACATCTCTGATCTCGGATCTGCACGGGTACACGCGCGGAGACACCATGAC	1186
QY	966	CTTCTCGGACGGGCTGACCTGAACCGGACCCAGATGCACAAACGGTGGCTTCGGCCCTT	1025
DB	1187	CTTCTCGGACGGGCTGACCTGAACCGGACCCAGATGCACAAACGGTGGCTTCGGCCCTT	1246
QY	1026	CACGACCTGTCTTTGCTTCGCCCAACAGATGCTGCCCTTGGAGATGGATGATCGGA	1085
DB	1247	CACGACCTGTCTTTGCTTCGCCCAACAGATGCTGCCCTTGGAGATGGATGATCGGA	1306
QY	1086	GACGGGCTGCTCAGCGGCATCTGCCTCATCTGCGGAGAGCCGAGGACCTGGAGAGCC	1145
DB	1307	GACGGGCTGCTCAGCGGCATCTGCCTCATCTGCGGAGAGCCGAGGACCTGGAGAGCC	1366
QY	1146	GGACCGGCTGACATGCTGCAGGAGCGCTGCTGCAGGCGCTTAAAGGTTCTAGCTCGGAA	1205
DB	1367	GGACCGGCTGACATGCTGCAGGAGCGCTGCTGCAGGCGCTTAAAGGTTCTAGCTCGGAA	1426
QY	1206	CGGAGGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGAAGATTACTGACCTGG	1265
DB	1427	CGGAGGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGAAGATTACTGACCTGG	1486
QY	1266	AAGCATCAGCCCAAGGGGCTGAGCGGCTGATCAGCTGAAGATGGAGATCCCGGGCTC	1325
DB	1487	AAGCATCAGCCCAAGGGGCTGAGCGGCTGATCAGCTGAAGATGGAGATCCCGGGCTC	1546
QY	1326	CATCCCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGCGCTGGACACTCTCAGCGG	1385
DB	1547	CATCCCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGCGCTGGACACTCTCAGCGG	1606
QY	1386	ACAGCCGGGGGTGGGGGCGGGGACGGGGTGGCTGCCCCCGCCCGCAGGAGCTGTAG	1445
DB	1607	ACAGCCGGGGGTGGGGGCGGGGACGGGGTGGCTGCCCCCGCCCGCAGGAGCTGTAG	1665
QY	1446	CCCCAGCCTCAGCCCCAGCTCCAAAGAGCAGCCGCGCCACCCACTCCCCGTGACCCGCC	1505
DB	1666	CCCCAGCCTCAGCCCCAGCTCCAAAGAGCAGCCGCGCCACCCACTCCCCGTGACCCGCC	1725
QY	1506	CAGCCACATGGACACAGCCCTCGCCCTCGCCCGCGGCTTTTCTGCTGCTTTCTACCGAC	1565
DB	1726	CAGCCACATGGACACAGCCCTCGCCCTCGCCCGCGGCTTTTCTGCTGCTTTCTACCGAC	1785
QY	1566	CATGTGACCCCGCAGCCCTGCCCCACCTGCGCTCCCGGGCAGTACTGGGACCTTC	1625
DB	1786	CATGTGACCCCGCAGCCCTGCCCCACCTGCGCTCCCGGGCAGTACTGGGACCTTC	1845
QY	1626	CCTGGGGAGCGGGGAGGAGGAGCAGCTCCTTTGGACAGAGGCTTGGCCCTCTAGTG	1685
DB	1846	CCTGGGGAGCGGGGAGGAGGAGCAGCTCCTTTGGACAGAGGCTTGGCCCTCTAGTG	1905
QY	1686	GACTGCTGCTCCACAGCCTGGGCTGACOTCAGAGGCGGAGCCAGGAATGAGTAGG	1745
DB	1906	GACTGCTGCTCCACAGCCTGGGCTGACOTCAGAGGCGGAGCCAGGAATGAGTAGG	1965
QY	1746	CCCTGCTGCTGGGCTCAGGATGGGTTCCTGGGGCTCTGTTTCATCAAGACACCCCTC	1805
DB	1966	CCCTGCTGCTGGGCTCAGGATGGGTTCCTGGGGCTCTGTTTCATCAAGACACCCCTC	2025

```
QY 1806 TGCCAGCTCACCACATCTTCATCACCAGCAAGCCGAGGACTGGCTCCCCCATCTCA 1865
Db 2026 TGCCAGCTCACCACATCTTCATCACCAGCAAGCCGAGGACTGGCTCCCCCATCTCA 2085
QY 1866 GAACACTACAAGCCATGTCTCCCGAGCTGGGGAACCTCAACCTCCCTGTACATACCTGCCATACCAAC 1985
Db 2086 GAACACTACAAGCCATGTCTCCCGAGCTGGGGAACCTCAACCTCCCTGTACATACCTGCCATACCAAC 2205
QY 1926 TGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTACATACCTGCCATACCAAC 1985
Db 2146 TGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTACATACCTGCCATACCAAC 2205
QY 1986 CCAGGTATTAAATCTCGCTGGTTCCTTTTATTTTAAATTTTGTGTTGATTTTTT 2045
Db 2206 CCAGGTATTAAATCTCGCTGGTTCCTTTTATTTTAAATTTTGTGTTGATTTTTT 2265
QY 2046 AATAAGATTTTCATTTTAAGCAACAAAAA 2080
Db 2266 AATAAGATTTTCATTTTAAGCAACAAAAA 2300

RESULT 2
BC008727 2441 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, retinoic acid receptor, alpha, clone MGC:1651
DEFINITION IMAGE:3163891, mRNA, complete cds.
ACCESSION BC008727
VERSION BC008727.1 GI:14250549
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2441)
Direct Submission
Submitted (25-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs@email.nih.gov
Tissue Procurement: ARCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Shenchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McLoskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 5 Row: 9 Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4506418.
FEATURES
Location/Qualifiers
1..2441
/organism="Homo sapiens"
/db_xref="LocusID:5914"
/db_xref="taxon:9606"
/clone="MGC:1651 IMAGE:3163891"
/tissue_type="Brain, neuroblastoma"
/clone_lib="NIH_MGC_19"
```

CDS

```
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
465..1853
/codon_start=1
/product="retinoic acid receptor, alpha"
/protein_id="AAH08727.1"
/db_xref="GI:14250550"
/translation="MAGNSSSSCTPTGGGHLNGYVPVYAFFPPLMGLGSLPPGALTTT
OHOLPVSGYSTPSPATITETQSSSEIEVPSPPPLPRIVKPCVCDKSSGYHGV
SACGCKGFFRRSRIOKNVYVTCRDKNCLINKVNRRCOYCRLOKCFVGMKSESVRN
DRNKKKEVPKPESESTLTPEVGLIEKVRKKAHOETFPALCQKGYTNNSSQVRV
SLDLDLWDFKSELSTKCIITKTFEFAKQLPFTLTIAQDITLLKAAACDLILIRICTR
YTPQDWTFTSDGLTLNARTOMHNAFGFLDTLVFAFANQLPLENDAETGLLSAICL
CGRDQLEQPDRLVMDLQEPLEALKVYVRKRRSRPHMFPMKMKITDRLRISAKGA
EVRTTLKMEIPGSMPLLIQEMLENSGLDTLSGPPGGGRDGLLAPPGSCPSLSLP
SSNRPSPATHSP"
BASE COUNT 510 a 822 c 672 g 437 t
ORIGIN
Query Match 85.5%; Score 1783.2; DB 9; Length 2441;
Best Local Similarity 99.8%; Pred. No. 4.3e-313;
Matches 1796; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 288 CTCCATTGAGACCCAGACGAGCTTCTGAAGATAGTGCCAGCCCTCCCTCGCCACC 347
Db 641 CACCATTGAGACCCAGACGAGCTTCTGAAGATAGTGCCAGCCCTCCCTCGCCACC 700
QY 348 CCTCTACCCCGCATCTACAAGCCTTGTCTGTCTAGGACAAGTCTCAGGCTACCA 407
Db 701 CCTCTACCCCGCATCTACAAGCCTTGTCTGTCTAGGACAAGTCTCAGGCTACCA 760
QY 408 CTATGGGCTGACGCGCTGTGAGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGAGAA 467
Db 761 CTATGGGCTGACGCGCTGTGAGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGAGAA 820
QY 468 CATGTGTACACGCTGTACACGGGACAAAGTGCATCATCAACAAGTGCACCGGAACCC 527
Db 821 CATGTGTACACGCTGTACACGGGACAAAGTGCATCATCAACAAGTGCACCGGAACCC 880
QY 528 CTGCCAGTACTGCCGACTGCAGAAAGTGTCTTGAAGTGGGATGTCCAAGAGTCTGTAG 587
Db 881 CTGCCAGTACTGCCGACTGCAGAAAGTGTCTTGAAGTGGGATGTCCAAGAGTCTGTAG 940
QY 588 AAACGACCGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 647
Db 941 AAACGACCGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1000
QY 648 GCTGACGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAGGCGCAGCAGGAAACCTT 707
Db 1001 GCTGACGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAGGCGCAGCAGGAAACCTT 1060
QY 708 CCTGCGCTCTGCCAGTGGGCAATATACACTACGAACACAGCTCAGAACACAGCTGTCTC 767
Db 1061 CCTGCGCTCTGCCAGTGGGCAATATACACTACGAACACAGCTCAGAACACAGCTGTCTC 1120
QY 768 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTTCCACCAAGTGCATTAAGAC 827
Db 1121 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTTCCACCAAGTGCATTAAGAC 1180
QY 828 TGTGGAGTTCGCAAGAGCTGCCCGGCTTACCACCTTCCACCTCCGCGACAGATCATC 887
Db 1181 TGTGGAGTTCGCAAGAGCTGCCCGGCTTACCACCTTCCACCTCCGCGACAGATCATC 1240
QY 888 CCTCTCAAGGCTCGCTCGGATCATCTGTATCTCGGATCTGCACCGGCTACAGCC 947
Db 1241 CCTCTCAAGGCTCGCTCGGATCATCTGTATCTCGGATCTGCACCGGCTACAGCC 1300
QY 948 CGAGCAGGACACCATGACCTTCTCGGAGCGGCTGACCTTGAACCGGACCCAGATGACAA 1007
Db 1301 CGAGCAGGACACCATGACCTTCTCGGAGCGGCTGACCTTGAACCGGACCCAGATGACAA 1360
QY 1008 CGTGGCTTCGGCCCTTCACGAGCTGTGTCTTGCCTTCGCAACAGCTGCTGCCCT 1067
```

Db 1361 CGCTGGCTTGGCCCTCCACGACCTGGTCTTTGCCCTTCCGCAACCAAGCTCTGCCCCCT 1420
Qy 1068 GGAGATGATGATCGGAGACGGGCTGCTCAGCGCCATCTCCCTCATCTCGGAGACGG 1127
Db 1421 GGAGATGATGATCGGAGACGGGCTGCTCAGCGCCATCTCCCTCATCTCGGAGACGG 1480
Qy 1128 CCAGACCTGGAGACGCGGACCGGCTGACATGCTCAGGAGCGGCTGCTCGAGGCGCT 1187
Db 1481 CCAGACCTGGAGACGCGGACCGGCTGACATGCTCAGGAGCGGCTGCTCGAGGCGCT 1540
Qy 1188 AAGGTCTACCTGGGAGCGGAGCGGCTGACATGCTCAGGAGCGGCTGCTCGAGGCGCT 1247
Db 1541 AAGGTCTACCTGGGAGCGGAGCGGCTGACATGCTCAGGAGCGGCTGCTCGAGGCGCT 1600
Qy 1248 GAAGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGACGGGCTGATCACCTGAA 1307
Db 1601 GAAGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGACGGGCTGATCACCTGAA 1660
Qy 1308 GATGGAGATCCGGGCTCCATGCCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGG 1367
Db 1661 GATGGAGATCCGGGCTCCATGCCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGG 1720
Qy 1368 CCTGGACACTCTGAGCGGACAGCGGGGGTGGGGGGGAGCGGGGTGGGCTGCCGCC 1427
Db 1721 CCTGGACACTCTGAGCGGACAGCGGGGGTGGGGGGGAGCGGGGTGGGCTGCCGCC 1780
Qy 1428 CCCGCCAGGAGCTGTAGCGCCAGCCTCAGCGCCAGCTCCACAGACGAGCGCGGCCAC 1487
Db 1781 CCCGCCAGGAGCTGTAGCGCCAGCCTCAGCGCCAGCTCCACAGACGAGCGCGGCCAC 1840
Qy 1488 CCACTCCCGGTGACCGCCAGCCTGACACAGCGCCCTCGCCCTCGCGCCCGGCTTTT 1547
Db 1841 CCACTCCCGGTGACCGCCAGCCTGACACAGCGCCCTCGCCCTCGCGCCCGGCTTTT 1900
Qy 1548 CTCGTGCTTTTACGGACCATGTGACCGCCAGCAGCGCCCTCGCCCTCGCGCCCGG 1607
Db 1901 CTCGTGCTTTTACGGACCATGTGACCGCCAGCAGCGCCCTCGCCCTCGCGCCCGG 1960
Qy 1608 GCAGTACTGGGACCTTCCCTGGGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1667
Db 1961 GCAGTACTGGGACCTTCCCTGGGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2020
Qy 1668 AGGCTTGGGCTCAGTGGACTGCTGCTCCACAGCGCTGGGCTGACGCTCAGAGCGCGAG 1727
Db 2021 AGGCTTGGGCTCAGTGGACTGCTGCTCCACAGCGCTGGGCTGACGCTCAGAGCGCGAG 2080
Qy 1728 GCGAGGAATCAGTGGGCTGCTGCTGGGCTCAGGATGGGCTCGGGGGCTCGTG 1787
Db 2081 GCGAGGAATCAGTGGGCTGCTGCTGGGCTCAGGATGGGCTCGGGGGCTCGTG 2140
Qy 1788 TTCATCAGACACCCCTGCTGCCAGCTCACCACATCTTTCATCACCAGCAAGCGCCAGAC 1847
Db 2141 TTCATCAGACACCCCTGCTGCCAGCTCACCACATCTTTCATCACCAGCAAGCGCCAGAC 2200
Qy 1848 TTGGCTCCCGCATCCTCAGAACTCACAAGCCTGCTCCCGAGCTGGGGAACCTCAACCT 1907
Db 2201 TTGGCTCCCGCATCCTCAGAACTCACAAGCCTGCTCCCGAGCTGGGGAACCTCAACCT 2260
Qy 1908 CCCCGCTGCTGCTGGTGGACAGAGGGGTGGGACAGGGGCGGGGGTTCGCCCTGTAC 1967
Db 2261 CCCCGCTGCTGCTGGTGGACAGAGGGGTGGGACAGGGGCGGGGGTTCGCCCTGTAC 2320
Qy 1968 ATACCCCTGCCATACCAACCCAGGTATTAATCTCGCTGGTGTGTTTTTAATTTTAA-TT 2026
Db 2321 ATACCCCTGCCATACCAACCCAGGTATTAATCTCGCTGGTGTGTTTTTAATTTTAA-TT 2380
Qy 2027 TTTTGTGTTTTGATTTTTTAAAGAAATTTTCATTTTAAAGCACAATAAAAAAAAAAAAA 2086
Db 2381 TTTTGTGTTTTGATTTTTTAAAGAAATTTTCATTTTAAAGCACAATAAAAAAAAAAAAA 2440

RESULT 3
AR061121

LOCUS: AR061121 3036 bp DNA linear PAT 29-SEP-1999
DEFINITION: Sequence 1 from patent US 5843642.
ACCESSION: AR061121
VERSION: AR061121.1 GI:5988812
KEYWORDS: Unknown.
SOURCE: Unknown.
ORGANISM: Unclassified.
REFERENCE: 1 (bases 1 to 3036)
AUTHORS: Dmitrovsky, E., Warrell, R. P. Jr., Miller, W. H. Jr. and Frankel, S.
TITLE: Methods for detection of acute promyelocytic leukemia (APL)
JOURNAL: Patent: US 5843642-A 1 01-DEC-1998;
FEATURES: Location/Qualifiers
1..3036
BASE COUNT: 605 a 1018 c 890 g 523 t
ORIGIN: /organism="unknown"

Query Match 85.4%; Score 1782.2; DB 6; Length 3036;
Best Local Similarity 99.8%; Pred. No. 7.2e-313;
Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 290 CCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTGCACGCTTCCCTCGCCACCCC 349
Db 1250 CCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTGCACGCTTCCCTCGCCACCCC 1309
Qy 350 CTCCTACCCGGCATCTACAAGCCTTGGTTGTCTGTGTCAGACAAAGTCTCTAGGCTACCACT 409
Db 1310 CTCCTACCCGGCATCTACAAGCCTTGGTTGTCTGTGTCAGACAAAGTCTCTAGGCTACCACT 1369
Qy 410 ATGGGGTTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCGCGCCGACGATCCAGAAAGACA 469
Db 1370 ATGGGGTTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCGCGCCGACGATCCAGAAAGACA 1429
Qy 470 TGGTGTACACGTGTCAACGGGACAAAGACTGCATCATCAACAGGTGACCCGGAACCCCT 529
Db 1430 TGGTGTACACGTGTCAACGGGACAAAGACTGCATCATCAACAGGTGACCCGGAACCCCT 1489
Qy 530 GCCAGTACTGCCAGCTGTCAGAAAGTCTTGAAGTGGGCTGTCCAAAGGAGTCTGTGAGAA 589
Db 1490 GCCAGTACTGCCAGCTGTCAGAAAGTCTTGAAGTGGGCTGTCCAAAGGAGTCTGTGAGAA 1549
Qy 590 AGCAGCGAAACAAAGAGAGAGGTGCCCAAGCCGAGTCTCTGAGAGCTACACGC 649
Db 1550 AGCAGCGAAACAAAGAGAGAGGTGCCCAAGCCGAGTCTCTGAGAGCTACACGC 1609
Qy 650 TGACCGCGAGGTGGGGAGCTCATTTGAAAGTGGCGAAAGCGCACACAGGAAACCTTCC 709
Db 1610 TGACCGCGAGGTGGGGAGCTCATTTGAAAGTGGCGAAAGCGCACACAGGAAACCTTCC 1669
Qy 710 CTGCGCTCTGCCAGCTGGGCAATACACTACGACAAACAGCTCAGAACACAGTGTCTCTC 769
Db 1670 CTGCGCTCTGCCAGCTGGGCAATACACTACGACAAACAGCTCAGAACACAGTGTCTCTC 1729
Qy 770 TGGACATTTGACCTCTGGGCAATGTTCACTGTAAGTCTCTCCACAGTGCATCATTAAGACTG 829
Db 1730 TGGACATTTGACCTCTGGGCAATGTTCACTGTAAGTCTCTCCACAGTGCATCATTAAGACTG 1789
Qy 830 TGGAGTTGCCCAAGCAGCTGCCCGCTTCCACACCTTCCACCTCGCCGACAGATCACCC 889
Db 1790 TGGAGTTGCCCAAGCAGCTGCCCGCTTCCACACCTTCCACCTCGCCGACAGATCACCC 1849
Qy 890 TCCTCAAGGCTGCCCTGGGACATCTGATCTCTGGGATCTGTCAGCGGGTACACGCCCG 949
Db 1850 TCCTCAAGGCTGCCCTGGGACATCTGATCTCTGGGATCTGTCAGCGGGTACACGCCCG 1909
Qy 950 AGCAGGACACCATGACCTTCTCGGAGGGCTGACCTTGAACCGGACCCAGATGCACAACG 1009
Db 1910 AGCAGGACACCATGACCTTCTCGGAGGGCTGACCTTGAACCGGACCCAGATGCACAACG 1969
Qy 1010 CTGGCTTGGGCCCCCTCACCGACCTGGTGTGCTTTGCTTTCGCCCAACAGCTGCTGCCCTGG 1069
Db 1970 CTGGCTTGGGCCCCCTCACCGACCTGGTGTGCTTTGCTTTCGCCCAACAGCTGCTGCCCTGG 2029

QY	1070	AGATGGATGATCGGAGACGGGCTCTCAGCGCCATCTGCCTCATCTCGGAGACGCC	1129
Db	2030	AGATGGATGATCGGAGACGGGCTCTCAGCGCCATCTGCCTCATCTCGGAGACGCC	2089
QY	1130	AGGACTGAGCAGCGGACCGGCTGACATGCTCAGAGACCGCTCTCGAGCGCTAA	1189
Db	2090	AGGACTGAGCAGCGGACCGGCTGACATGCTCAGAGACCGCTCTCGAGCGCTAA	2149
QY	1190	AGGTTACTGCGGAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGA	1249
Db	2150	AGGTTACTGCGGAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGA	2209
QY	1250	AGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGGTGCATCACCTGNAGA	1309
Db	2210	AGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGGTGCATCACCTGNAGA	2269
QY	1310	TGAGATCCCGGCTCCATGCGGCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC	1369
Db	2270	TGAGATCCCGGCTCCATGCGGCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC	2329
QY	1370	TGGACACTGTAGCGGACAGCGGGGGTGGGGGGGGGAGCGGGGTGGCTTGCCTCC	1429
Db	2330	TGGACACTGTAGCGGACAGCGGGGGTGGGGGGGGGAGCGGGGTGGCTTGCCTCC	2389
QY	1430	CGCAGGACAGCTGTAGCGGACAGCGGGGGTGGGGGGGGGAGCGGGGTGGCTTGCCTCC	1489
Db	2390	CGCAGGACAGCTGTAGCGGACAGCGGGGGTGGGGGGGGGAGCGGGGTGGCTTGCCTCC	2449
QY	1490	ACTCCCGTACCGGACAGCGGGGGTGGGGGGGGGAGCGGGGTGGCTTGCCTCC	1549
Db	2450	ACTCCCGTACCGGACAGCGGGGGTGGGGGGGGGAGCGGGGTGGCTTGCCTCC	2509
QY	1550	CTGCGCTTTCTACCGACATGTGACCGCGACAGCGCTTGCCTCCCGGCTCCCGGGC	1609
Db	2510	CTGCGCTTTCTACCGACATGTGACCGCGACAGCGCTTGCCTCCCGGCTCCCGGGC	2569
QY	1610	AGTACTGGGACCTTCCCTGGGGAGCGGGGAGGAGGAGCGAGCTCTTGGACAGAG	1669
Db	2570	AGTACTGGGACCTTCCCTGGGGAGCGGGGAGGAGGAGCGAGCTCTTGGACAGAG	2629
QY	1670	GCTGGGCGCTCAGTGGAGCTGCTGCTCCACAGCGCTGGGCTCAGCTCAGGCGCGAGGC	1729
Db	2630	GCTGGGCGCTCAGTGGAGCTGCTGCTCCACAGCGCTGGGCTCAGCTCAGGCGCGAGGC	2689
QY	1730	CAGGAATGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1789
Db	2690	CAGGAATGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2749
QY	1790	CATCAAGACACCCCTCTGCCAGCTCACCACATCTTCAATCACCAGCAAAAGCGGAGCTT	1849
Db	2750	CATCAAGACACCCCTCTGCCAGCTCACCACATCTTCAATCACCAGCAAAAGCGGAGCTT	2809
QY	1850	GGTCCCGCATCTCAGAACTCAGAGCCATCTGCTCCCGAGCTGGGAACTCAACCTCC	1909
Db	2810	GGTCCCGCATCTCAGAACTCAGAGCCATCTGCTCCCGAGCTGGGAACTCAACCTCC	2869
QY	1910	CCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1969
Db	2870	CCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2929
QY	1970	ACCTGCCATACCAACCCAGGATTAATTTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2029
Db	2930	ACCTGCCATACCAACCCAGGATTAATTTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2989
QY	2030	TGCTTTTGAATTTTAAATGAATTTTCAATTTTAAAGCAAAAAA	2076
Db	2990	TGCTTTTGAATTTTAAATGAATTTTCAATTTTAAAGCAAAAAA	3036
RESULT 4			
196215			
LOCUS			

DEFINITION	Sequence 52 from patent US 5734039.		
ACCESSION	196215		
VERSION	196215.1		
KEYWORDS	GI:3940685		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 3036)		
AUTHORS	Calabretta B. and Skorski, T.		
TITLE	Antisense oligonucleotides targeting cooperating oncogenes		
JOURNAL	Patent: US 5734039-A 52 31-MAR-1998;		
FEATURES	Location/Qualifiers		
source	1..3036		
BASE COUNT	605 a	1018 c	890 g 523 t
ORIGIN	/organism="unknown"		
Query Match	85.4%; Score 1782.2; DB 6; Length 3036;		
Best Local Similarity	99.8%; Pred. No. 7.2e-313;		
Matches 1784; Conservative	0; Mismatches 3; Indels 0; Gaps 0;		
QY	290	CCATTGAGACCCAGAGCAGCAGTCTCTGAAGAGATAGTCCCGCAGCCCTCCTCGCCACCCC	349
Db	1250	CCATTGAGACCCAGAGCAGCAGTCTCTGAAGAGATAGTCCCGCAGCCCTCCTCGCCACCCC	1309
QY	350	CTCTACCCCGCATCTACAAGCCTTGTCTGTCTGAGGACAACTCTCAGCTACCACT	409
Db	1310	CTCTACCCCGCATCTACAAGCCTTGTCTGTCTGAGGACAACTCTCAGCTACCACT	1369
QY	410	ATGGGGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGAAAGACA	469
Db	1370	ATGGGGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGAAAGACA	1429
QY	470	TGTTGTACAGTGTACCGGGACAAAGAACTGCATCATCAACAAGGTGACCCGGAAACCCCT	529
Db	1430	TGTTGTACAGTGTACCGGGACAAAGAACTGCATCATCAACAAGGTGACCCGGAAACCCCT	1489
QY	530	GCCAGTACTGCGGACTGCGAAGTGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAGNA	589
Db	1490	GCCAGTACTGCGGACTGCGAAGTGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAGNA	1549
QY	590	ACGACCGAAACAAAGAAAGAGGAGTGCCCAAGCCGAGTGTCTGTGAGAGCTACACGC	649
Db	1550	ACGACCGAAACAAAGAAAGAGGAGTGCCCAAGCCGAGTGTCTGTGAGAGCTACACGC	1609
QY	650	TGACCGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAGCGCAGCAGGAAACCTTCC	709
Db	1610	TGACCGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAGCGCAGCAGGAAACCTTCC	1669
QY	710	CTGCGCTCTGCGAGCTGGGCAAAATACACTAGCAACAGCTCAGAACAACTGTCTCTC	769
Db	1670	CTGCGCTCTGCGAGCTGGGCAAAATACACTAGCAACAGCTCAGAACAACTGTCTCTC	1729
QY	770	TGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTTCCACCAAGTGCATCATTAAGACTG	829
Db	1730	TGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTTCCACCAAGTGCATCATTAAGACTG	1789
QY	830	TGGAGTTGCGCAAGCAGCTGCGCGCTTCAACACCTCACCCTCCCGACCAAGATCACCC	889
Db	1790	TGGAGTTGCGCAAGCAGCTGCGCGCTTCAACACCTCACCCTCCCGACCAAGATCACCC	1849
QY	890	TCCTCAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	949
Db	1850	TCCTCAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1909
QY	950	ACGAGGACACCATGACCTTCTCGGACGGCTGACCTCAACCGGACCCAGATGCACAGC	1009
Db	1910	ACGAGGACACCATGACCTTCTCGGACGGCTGACCTCAACCGGACCCAGATGCACAGC	1969
QY	1010	CTGGCTTGGCGCCCTCACCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1069
Db	1970	CTGGCTTGGCGCCCTCACCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2029

QY 1070 AGATGGATGATCGGAGACGGGGCTGCTCAGGCGCATCTGCTCATCTGCGGAGACGCC 1129
 Db 2030 AGATGGATGATCGGAGACGGGGCTGCTCAGGCGCATCTGCTCATCTGCGGAGACGCC 2089
 QY 1130 AGGACCTGGAGACGGCGGCTGGACATGCTGTCAGGACCGCTGCTGAGGCGCTAA 1189
 Db 2090 AGGACCTGGAGACGGCGGCTGGACATGCTGTCAGGACCGCTGCTGAGGCGCTAA 2149
 QY 1190 AGGTCTACGTGCGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAATGA 1249
 Db 2150 AGGTCTACGTGCGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAATGA 2209
 QY 1250 AGATTACTGACTCGGAAGCATACGCGCAAGGGGGCTGAGCGGGTGATCAGCTGAAGA 1309
 Db 2210 AGATTACTGACTCGGAAGCATACGCGCAAGGGGGCTGAGCGGGTGATCAGCTGAAGA 2269
 QY 1310 TGGAGATCCCGGGCTCCATGCGCGCTCTCATCCAGGAATGTTGGAGACTCAGAGGCC 1369
 Db 2270 TGGAGATCCCGGGCTCCATGCGCGCTCTCATCCAGGAATGTTGGAGACTCAGAGGCC 2329
 QY 1370 TGGACACTCTGAGCGGACAGCGGGGGGTGGGGGGCGGGGGTGGCTGCCGCCGCC 1429
 Db 2330 TGGACACTCTGAGCGGACAGCGGGGGGTGGGGGGCGGGGGTGGCTGCCGCCGCC 2389
 QY 1430 CGCCAGGAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCCACAGAGAGCGCCGCCAGCC 1489
 Db 2390 CGCCAGGAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCCACAGAGAGCGCCGCCAGCC 2449
 QY 1490 ACTCCCGTGTAGCCCGCAGCTCAGCCCGCAGCTCAGCCCGCAGCTCAGCCCGCAGCT 1549
 Db 2450 ACTCCCGTGTAGCCCGCAGCTCAGCCCGCAGCTCAGCCCGCAGCTCAGCCCGCAGCT 2509
 QY 1550 CTGCGTTTCTACCGACCATGTGACCCCGCAGCTCAGCCCGCAGCTCAGCCCGCAGCT 1609
 Db 2510 CTGCGTTTCTACCGACCATGTGACCCCGCAGCTCAGCCCGCAGCTCAGCCCGCAGCT 2569
 QY 1610 AGTACTGGGACCTTCCCTGGGGACCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1669
 Db 2570 AGTACTGGGACCTTCCCTGGGGACCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2629
 QY 1670 GCTTGGGCGCTCAGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1729
 Db 2630 GCTTGGGCGCTCAGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2689
 QY 1730 CAGGAATGAGTGAGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789
 Db 2690 CAGGAATGAGTGAGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2749
 QY 1790 CATCAGACACCCCTCTGCGCCAGCTCAGCCATCTTCCATCAGCAGCAAGCCAGGACTT 1849
 Db 2750 CATCAGACACCCCTCTGCGCCAGCTCAGCCATCTTCCATCAGCAGCAAGCCAGGACTT 2809
 QY 1850 GGTCTCCCGATCTCAGAACTCAGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1909
 Db 2810 GGTCTCCCGATCTCAGAACTCAGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2869
 QY 1910 CCCTGCTCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1969
 Db 2870 CCCTGCTCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2929
 QY 1970 ACCCTGCCATACCAACCCAGGATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2029
 Db 2930 ACCCTGCCATACCAACCCAGGATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2989
 QY 2030 TTGTTTTGATTTTTTAATAAGATTTTCAATTTTAAGCAAAAAA 2076
 Db 2990 TTGTTTTGATTTTTTAATAAGATTTTCAATTTTAAGCAAAAAA 3036

RESULT 5

HUMPLRAR

LOCUS

DEFINITION

Human PML-RAR protein (PML-RAR) mRNA, complete cds.

3036 bp mRNA linear PRI 08-JAN-1995

ACCESSION

M73779.1

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

FEATURES

source

gene

CDS

Location/Qualifiers

1. 3036

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="chromosome t(15;17)"

/cell_type="bone marrow"

1. 3036

/gene="PML-RAR"

67. 2460

/gene="PML-RAR"

/codon_start=1

/product="PML-RAR protein"

/protein_id="AAA60126.1"

/db_xref="GI:190125"

/translation="MEPAPARSPRPODPAPOPTMPPTPSEGRQSPSPSPPTER

APASEFEFLRCQCOAEAKPKLLCLHTLCSGLEASGMCQPCICQAPNPLGADTP

ALDNVFESLQRLSVYRIQVDAQVTRCKESADFCFECEQLLCAKCFPAHOMFLK

HEARPLAELRNQSVREFLDGTRKTNNIFCSNPNNHTPLTSLIYCRGSKPLCCSCALL

DSSHSELKCDISAEIQORQELDAMTQALQDSAFGAVHQAHAHQVQLGARAETE

ELJRVRVQVAVHVRQERELLDAVDARYODYEMASRLRLDVLQRIITGSAALVO

RMKYASDQEVLDHGLFLQALFLROEPQSLQAAVDTGDFDEKVLQDLSSCITO

KGALETQSSSEELVPSPPPLPRYIKFCVQDQSSGTHYGVACEGCKGFFRRS

IQKNMYTCHRDKNILNKVRNRCQYLRCKFEVMSKESVENDRNKKKKEVPKPE

CSBSYTLTPEVGLIEIKVRAHOETFPALCOLGKYTTNNSEQRVSLDIDLWDFSEL

STKCIITKTEFAKFGFTLTADQITLLKACLDLILIRICTRYTPEQDTMTFSDG

LTNRTQMHNAGFPLTDLVAFANOLLPLEMDDAETGLLSAICLCIGDRDLQSPDR

VMDQPLLEALKAVYVRKRPSPHMPKMLKTTDLRSISAKGAERVTLKMEIPGS

MPPLIOEMLENSGLDTLSQPGGGGRDGLAPPQSCSPSLSPSSNRSPSPATHSP"

BASE COUNT 505 a 1018 c 890 g 523 t

ORIGIN chromosome t(15;17) translocation.

Query Match 85.4%; Score 1782.2; DB 9; Length 3036;

Best Local Similarity 99.8%; Pred. No. 7.2e-313;

Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 290 CCATTGAGACCCAGAGCAGCAGTTCCTGAAGAGATAGTGCCAGCCCTCCCTCGCACCCC 349

Db 1250 CCATTGAGACCCAGAGCAGCAGTTCCTGAAGAGATAGTGCCAGCCCTCCCTCGCACCCC 1309

QY 350 CTCTACCCCGCATCTACAAGCCTTGCTTCTCTCAGGACAACTCTCAGGCTACCACT 409

Db 1310 CTCTACCCCGCATCTACAAGCCTTGCTTCTCTCAGGACAACTCTCAGGCTACCACT 1369

QY 410 ATGGGGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCCCGCGCAGCATCCAGAGACA 469

Db 1370 ATGGGGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCCCGCGCAGCATCCAGAGACA 1429

QY 470 TGGTGTACAGTGTACCCGGACAAAGAACTGCATCATCAAGCTGACCCGGAAACCCCT 529

Db 1430 TGGTGTACAGTGTACCCGGACAAAGAACTGCATCATCAAGCTGACCCGGAAACCCCT 1489

QY 530 GCCAGTACTGCCGACTGCGAAGTGTCTTGAAGTGGGCATGTCCAAGAGTCTCTGAGAA 589

Db 1490 GCCAGTACTGCCGACTGCGAAGTGTCTTGAAGTGGGCATGTCCAAGAGTCTCTGAGAA 1549

QY 590 ACCAGCAACACAGAGAGAGAGTGCCCAAGCCCGAGTGTCTGAGAGCTACAGC 649

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||


```
Db 519 CTGCCAGTACTGCCGACTGCAGAGTGTCTTTGAAGTGGGCATGTCCAAGGAGTCTGTGAG 578
Qy 588 AAACCAACCAACACAGAGAAGAGAGGTGCCCGCCAGCCGAGTGTCTGAGAGCTACAC 647
Db 579 AAACCAACCAACAAAGAGAAGAGAGGTGCCCGCCAGCCGAGTGTCTGAGAGCTACAC 638
Qy 648 GCTGACGCCGAGGTGGGGAGCTCATTCAGAGAGTGGCCAAAGCGCACCGAAGAACCTT 707
Db 639 GCTGACGCCGAGGTGGGGAGCTCATTCAGAGAGTGGCCAAAGCGCACCGAAGAACCTT 698
Qy 708 CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAACAAACAGCTCAGAACAAACGTGTCTC 767
Db 699 CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAACAAACAGCTCAGAACAAACGTGTCTC 758
Qy 768 TCTGGACATTGACCTCTGGGACAAAGTTCAAGTAACTCTCCACCAAGTGCATCATTAAGAC 827
Db 759 TCTGGACATTGACCTCTGGGACAAAGTTCAAGTAACTCTCCACCAAGTGCATCATTAAGAC 818
Qy 828 TGTGGAGTTCCCAAGCAGCTGCCGGCTTACCACCTTACCACCTTACCACCTGCGCGACGATCAC 887
Db 819 TGTGGAGTTCCCAAGCAGCTGCCGGCTTACCACCTTACCACCTTACCACCTGCGCGACGATCAC 878
Qy 888 CCTCCTCAAGGCTGCCCTGSGACATCCTGATCTGCGGATCTGACGCGGATGCGACGCGGATCACGCC 947
Db 879 CCTCCTCAAGGCTGCCCTGSGACATCCTGATCTGCGGATCTGACGCGGATGCGACGCGGATCACGCC 938
Qy 948 CGAGCAGACACCAATGACCTTCTCGGAGCGGCTGACCTGAAACCGGACCCAGATGCACAA 1007
Db 939 CGAGCAGACACCAATGACCTTCTCGGAGCGGCTGACCTGAAACCGGACCCAGATGCACAA 998
Qy 1008 CGCTGGCTTCCGCCCTTCCAGCAGCTGGTCTTTGGCTTGGCCAAACAGCTGCTGCCCT 1067
Db 999 CGCTGGCTTCCGCCCTTCCAGCAGCTGGTCTTTGGCTTGGCCAAACAGCTGCTGCCCT 1058
Qy 1068 GGAGATGGATGATCGGAGAGCGGGCTGCTCAGCGCCATCTGCGCTCATCTCGGAGACCG 1127
Db 1059 GGAGATGGATGATCGGAGAGCGGGCTGCTCAGCGCCATCTGCGCTCATCTCGGAGACCG 1118
Qy 1128 CCAGGACCTGGAGACCGCGGAGCGGTGGACATGCTCAGGAGCGCTGCTGGAGGCGCT 1187
Db 1119 CCAGGACCTGGAGACCGCGGAGCGGTGGACATGCTCAGGAGCGCTGCTGGAGGCGCT 1178
Qy 1188 AAGGTCTAGTGGGAAGCGGAGCGCCAGCCGCCCCACATGTTCCCAAGATGCTAAT 1247
Db 1179 AAGGTCTAGTGGGAAGCGGAGCGCCAGCCGCCCCACATGTTCCCAAGATGCTAAT 1238
Qy 1248 GAAGATTACTGACCTGCGAGACATCAGCCCAAGGGGCTGAGCGGTGATCAGCTGAA 1307
Db 1239 GAAGATTACTGACCTGCGAGACATCAGCCCAAGGGGCTGAGCGGTGATCAGCTGAA 1298
Qy 1308 GATGGAGATCCCGGCTCCATGCCGCTCTCATCCAGGAATGTTGGAGAACTCAGAGG 1367
Db 1299 GATGGAGATCCCGGCTCCATGCCGCTCTCATCCAGGAATGTTGGAGAACTCAGAGG 1358
Qy 1368 CTTGGACATCTGAGCGGACAGCCGGGGGTGGGGGCGGAGCGGGGTGGCTGCCCTCC 1427
Db 1359 CTTGGACATCTGAGCGGACAGCCGGGGGTGGGGGCGGAGCGGGGTGGCTGCCCTCC 1418
Qy 1428 CCGCCAGCAGCTGTGAGCCCGCAGCTCAGCCCGCCAGCTTCCACAGAGAGAGCCCGCCAC 1487
Db 1419 CCGCCAGCAGCTGTGAGCCCGCAGCTCAGCCCGCCAGCTTCCACAGAGAGAGCCCGCCAC 1478
Qy 1488 CCATCTCCCGTACCGCCCGCCAGCTGACAGACAGCTTCCGCTCCGCTCCGCTCCGCTTTT 1547
Db 1479 CCATCTCCCGTACCGCCCGCCAGCTGACAGACAGCTTCCGCTCCGCTCCGCTCCGCTTTT 1538
Qy 1548 CTCTGCTTTTACCGACCATGTACCCCGCCAGCCCTGCCCTCCGCTCCGCTCCGCTTTT 1607
Db 1539 CTCTGCTTTTACCGACCATGTACCCCGCCAGCCCTGCCCTCCGCTCCGCTCCGCTTTT 1598
Qy 1608 CGAGTACTGGGACCTTCCCTGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667
Db 1599 CGAGTACTGGGACCTTCCCTGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658
```

```
Qy 1668 AGGCCTGGGCCCTCACTGGACTGCCTGCCTCCACAGCCTGGGCTGAGCTCAGAGGCCGAG 1727
Db 1659 AGGCCTGGGCCCTCACTGGACTGCCTGCCTCCACAGCCTGGGCTGAGCTCAGAGGCCGAG 1718
Qy 1728 GCCAGAACTGAGTGAAGGCCCTGCTGGGCTCCTCAGGATGGTCTCCTGGGGCTCGTG 1787
Db 1719 GCCAGAACTGAGTGAAGGCCCTGCTGGGCTCCTCAGGATGGTCTCCTGGGGCTCGTG 1778
Qy 1788 TTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGCCAGGAC 1847
Db 1779 TTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGCCAGGAC 1838
Qy 1848 TTGGCTCCGCCCTCCTCAGAACTCACAAGCCTTGTCTCCACCTGGGAGAACCTCAACCT 1907
Db 1839 TTGGCTCCGCCCTCCTCAGAACTCACAAGCCTTGTCTCCACCTGGGAGAACCTCAACCT 1898
Qy 1908 CCCCCCTGCCCTGGTGTGACAGAGGGGTGGACAGGGGGGGGGTTCCTCCCTGTAC 1967
Db 1899 CCCCCCTGCCCTGGTGTGACAGAGGGGTGGACAGGGGGGGGGTTCCTCCCTGTAC 1958
Qy 1968 ATACCTGCCATACCAACCCAGGTATTAAATTCCTCGCTGGTTCCTGTTTAAATTT 2027
Db 1959 ATACCTGCCATACCAACCCAGGTATTAAATTCCTCGCTGGTTCCTGTTTAAATTT 2018
Qy 2028 TTTTCTTTTGAATTTTAAATTAAGAAATTTTCATTTAAGCACAAAAAATAAAAA 2085
Db 2019 TTTTCTTTTGAATTTTAAATTAAGAAATTTTCATTTAAGCACATTTATTAAGGAA 2076
```

```
RESULT 7
HSRRA
LOCUS
DEFINITION Human mRNA for receptor of retinoic acid.
ACCESSION X06614
VERSION X06614.1 GI:36156
KEYWORDS hormone receptor; retinoic acid receptor.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2907)
AUTHORS Giguere,V., Ong,E.S., Segui,P. and Evans,R.M.
TITLE Identification of a receptor for the morphogen retinoic acid
JOURNAL Nature 330 (6149), 624-629 (1987)
MEDLINE 88065922
PUBMED 2825036
REFERENCE
2 (bases 1 to 2907)
AUTHORS Banihmad,A., Kohne,A.C. and Renkawitz,R.
TITLE A transferable silencing domain is present in the thyroid hormone
receptor, in the v-erbA oncogene product and in the retinoic acid
receptor
JOURNAL EMBO J. 11 (3), 1015-1023 (1992)
MEDLINE 92191988
PUBMED 1347744
FEATURES
Location/Qualifiers
1..2907
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="lambda hK1R"
/clone_lib="lambda gt10"
103..1491
/note="put. retinoic acid receptor (AA 1-462)"
/codon_start=1
/protein_id="CAA29829.1"
/db_xref="GI:36157"
/db_xref="SWISS-PROT:P10276"
CDS
/translation="MASNSSCTPGGHLNGYVPPYAFFFPMLGGLSPPGALTTL
QHLFVSGYSTPATIETQSSSSEIVSPSPPLPRYKPCFVCDKSSGVHYGV
SACECKGFRRSIQKNMVTCTHRDKNCILINKVRNRCOYCRLOKFEVGMSEVRN
DNKKKKVPEKPSSEYLTTFEVELIEKRAHQETFPALCOLQKTYTNNSEQRV
SLDIDLWKFSELSTKCIKTVEFAKQLPGFTLTADITDILKACLDLILRICTR
YTPEDTWTFTSDGLTLNRTQMHNAGFGPLTLDFAFANQLPLEMDAETGLSLACL
```

1119 CCAGGACCTGGAGCAGCCGGACCGGTGGACATGCTGCAGGAGCCGCTGCTCGAGGGCGCT 117

REFERENCE 1 (bases 1 to 2928)

ORGANISM
UNKNOWN.
Unclassified.

REFERENCE
1 (bases 1 to 2928)

AUTHORS Dmitrovsky, E., Warrell, R. P. Jr., Miller, W. H. Jr. and Frankel, S.
TITLE Methods for detection of acute promyelocytic leukemia (APL)
JOURNAL Patent: US 5843642-A 3 01-DEC-1998;
FEATURES Location/Qualifiers
source 1. .2928
BASE COUNT 582 a 1011 c 789 g 546 t
ORIGIN

Query Match		85.3%;	Score 1778.8;	DB 6;	Length 2928;
Best Local Similarity		99.3%;	Pred. No. 3e-312;		
Matches 1786;		Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;
QY	288	CTCATTTAGACCCAGAGCAGCAGTCTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC	347		
DB	279	CACCATTTAGACCCAGAGCAGCAGTCTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC	338		
QY	348	CCCTCTACCCCGCATCTACAGCCCTTGTCTGTCTCAGGACAAAGTCTCTCAGGCTACCA	407		
DB	339	CCCTCTACCCCGCATCTACAGCCCTTGTCTGTCTCAGGACAAAGTCTCTCAGGCTACCA	398		
QY	408	CTATGGGTGACGCGCTGTGAGGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGAGAA	467		
DB	399	CTATGGGTGACGCGCTGTGAGGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGAGAA	458		
QY	468	CATGGTGTACAGTGTCTACCGGGCAAGAACTGCATCATCAACAAGGTGACCCGGAAACC	527		
DB	459	CATGGTGTACAGTGTCTACCGGGCAAGAACTGCATCATCAACAAGGTGACCCGGAAACC	518		
QY	528	CTGCAGTACTCCGACTGCAGAAAGTCTTGAAGTGGGCATGTCCAAGAGTCTCTGTAG	587		
DB	519	CTGCAGTACTCCGACTGCAGAAAGTCTTGAAGTGGGCATGTCCAAGAGTCTCTGTAG	578		
QY	588	AAACGACCGAACAAGAGAGAGAGTGTCCCAAGAGTGCACCAAGCGCACCAAGACCTT	647		
DB	579	AAACGACCGAACAAGAGAGAGAGTGTCCCAAGAGTGCACCAAGCGCACCAAGACCTT	638		
QY	648	GCTGACGCGGAGTGGGGAGCTCATTTAGAAAGTGCACCAAGCGCACCAAGACCTT	707		
DB	639	GCTGACGCGGAGTGGGGAGCTCATTTAGAAAGTGCACCAAGCGCACCAAGACCTT	698		
QY	708	CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAACAACAGCTCAGAACACGTGTCTC	767		
DB	699	CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAACAACAGCTCAGAACACGTGTCTC	758		
QY	768	TCTGACATTCACCTCTGGGCAAGTCTCAGTCAACTCTCCACCAAGTGCATCATTAAGAC	827		
DB	759	TCTGACATTCACCTCTGGGCAAGTCTCAGTCAACTCTCCACCAAGTGCATCATTAAGAC	818		
QY	828	TGTGGAGTTCGCCAAGCAGCTGCCGGCTTCCACCACTCCACCACTCCGCGCAGATCAC	887		
DB	819	TGTGGAGTTCGCCAAGCAGCTGCCGGCTTCCACCACTCCACCACTCCGCGCAGATCAC	878		
QY	888	CCTCCTCAAGCTGCCCTGGAGATCCTGATCTCGGGATCTGCACCGCGGTACACGCC	947		
DB	879	CCTCCTCAAGCTGCCCTGGAGATCCTGATCTCGGGATCTGCACCGCGGTACACGCC	938		
QY	948	CGACGAGACACCATGACCTTCTCGGAGGGCTGACCTTGACCGGACCCAGATGCACAA	1007		
DB	939	CGACGAGACACCATGACCTTCTCGGAGGGCTGACCTTGACCGGACCCAGATGCACAA	998		
QY	1008	CGCTGGCTTCGGCCCTTCCACCGACCTGGTCTTTCGCTTTCGCCCAACAGCTGTGCCCT	1067		
DB	999	CGCTGGCTTCGGCCCTTCCACCGACCTGGTCTTTCGCTTTCGCCCAACAGCTGTGCCCT	1058		
QY	1068	GGAGATGGATGATCGGAGAGGGGCTGCTCAGCGCCATCTGCTCATCTCGGAGACCG	1127		
DB	1059	GGAGATGGATGATCGGAGAGGGGCTGCTCAGCGCCATCTGCTCATCTCTCGGAGACCG	1118		
QY	1128	CCAGGACCTGGAGACCGCGGCTGGACATCTCGAGGAGCGCTGTCTGGAGGGCT	1187		
DB	1119	CCAGGACCTGGAGACCGCGGCTGGACATCTCGAGGAGCGCTGTCTGGAGGGCT	1178		

QY	1188	AAAGTCTACCTGCGGAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAT	1247
DB	1179	AAAGTCTACCTGCGGAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAT	1238
QY	1248	GAAGATTACTGACCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGTACACCTGAA	1307
DB	1239	GAAGATTACTGACCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGTACACCTGAA	1298
QY	1308	GATGGAGATCCCGGCTCCATGCGCCCTTCTCATCAGGAAATGTTGGAACTCAGAGG	1367
DB	1299	GATGGAGATCCCGGCTCCATGCGCCCTTCTCATCAGGAAATGTTGGAACTCAGAGG	1358
QY	1368	CCTGGACACTCTGAGCGGACAGCGGGGGTGGGGGGGAGCGGGGTGGCTGCCGCC	1427
DB	1359	CCTGGACACTCTGAGCGGACAGCGGGGGTGGGGGGGAGCGGGGTGGCTGCCGCC	1418
QY	1428	CCGCCAGGAGCTGTAGCCCGCCAGCCTCAGCCCGAGCTTCCAAACAGACGACCGGCCAC	1487
DB	1419	CCGCCAGGAGCTGTAGCCCGCCAGCCTCAGCCCGAGCTTCCAAACAGACGACCGGCCAC	1478
QY	1488	CCACTCCCGTGTAGCCCGCCAGCCTCAGCCCGAGCTTCCAAACAGACGACCGGCCCTT	1547
DB	1479	CCACTCCCGTGTAGCCCGCCAGCCTCAGCCCGAGCTTCCAAACAGACGACCGGCCCTT	1538
QY	1548	CTCTGCTTCTTACCGACCATGTGACCCCGACCGACCGCTTCCCGCCCTCCCGG	1607
DB	1539	CTCTGCTTCTTACCGACCATGTGACCCCGACCGACCGCTTCCCGCCCTCCCGG	1598
QY	1608	GCACTACTGGGGACCTTCCCTGGGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGG	1667
DB	1599	GCACTACTGGGGACCTTCCCTGGGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGG	1658
QY	1668	AGSCCTGGGCGCTCAGTGGAGTGCCTGCCACAGCTGGGGTGCAGTGCAGAGGCGGAG	1727
DB	1659	AGSCCTGGGCGCTCAGTGGAGTGCCTGCCACAGCTGGGGTGCAGTGCAGAGGCGGAG	1718
QY	1728	GCCAGAACTGAGTGGGCGCTTGGCTTCTCAGTGGGTCTCAGTGGGTCTGGGGGCTCTG	1787
DB	1719	GCCAGAACTGAGTGGGCGCTTGGCTTCTCAGTGGGTCTCAGTGGGTCTGGGGGCTCTG	1778
QY	1788	TTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCAATCACCAGCAACGCCAGAC	1847
DB	1779	TTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCAATCACCAGCAACGCCAGAC	1838
QY	1848	TTGGCTCCCGCTCAGAACTCAGAACTCAGAACTTGTCTCCAGCTGGGGAACTCAACCT	1907
DB	1839	TTGGCTCCCGCTCAGAACTCAGAACTCAGAACTTGTCTCCAGCTGGGGAACTCAACCT	1898
QY	1908	CCCCCTGCTCGTGGTGGTGGAGAGGGGTGGAGAGGGGTGGAGAGGGGTGGGGGCTGAC	1967
DB	1899	CCCCCTGCTCGTGGTGGTGGAGAGGGGTGGAGAGGGGTGGAGAGGGGTGGGGGCTGAC	1958
QY	1968	ATACCTGTCATACCAACCCAGGATTAATCTCGCTGCTGTTTGTGTTTATTTTAAATTT	2027
DB	1959	ATACCTGTCATACCAACCCAGGATTAATCTCGCTGCTGTTTGTGTTTATTTTAAATTT	2018
QY	2028	TTTTGTTGATTTTTTAAAGAAATTTTCAATTTTAAAGCAAAAAAAGGAGGAGGAGGAGG	2085
DB	2019	TTTTGTTGATTTTTTAAAGAAATTTTCAATTTTAAAGCAAAAAAAGGAGGAGGAGGAGG	2076

RESULT 9

108117

LOCUS

DEFINITION

Sequence 1 from Patent EP 0325849.

ACCESSION

I08117

VERSION

I08117.1

KEYWORDS

GI:589170

Unknown.

ORGANISM

Unknown.

Unclassified.

REFERENCE

1 (bases 1 to 2908)

AUTHORS

Evans, R.M., Giguere, V., Ong, E.S., Segul, P.S., Umesono, K. and

DNA

2908 bp

linear

PAT 02-DEC-1994

AUTHORS Tsai, S. and Collins, S.J.
TITLE Creating novel hematopoietic cell lines by expressing altered retinoic acid receptors
JOURNAL Patent: US 5830760-A 1 03-NOV-1998;
FEATURES Location/Qualifiers
1. 2940
source /organism="unknown"
BASE COUNT 594 a 1012 c 788 g 546 t
ORIGIN

Query Match 85.2%; Score 1777.2; DB 6; Length 2940;
Best Local Similarity 99.3%; Pred. No. 6e-312;
Matches 1785; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 288 CTCATTGAGACCCAGACAGCAGTCTGAGAGAGATAGTCCCGACCCCTCCCTCCGCCACC 347
DB 279 CACCATTTGAGACCCAGACAGCAGTCTGAGAGAGATAGTCCCGACCCCTCCCTCCGCCACC 338
QY 348 CCCTCTACCCCGCATCTACAAAGCCTTGCTTTGTCTGTGAGGACAAAGTCTCTCAGGCTACCA 407
DB 339 CCCTCTACCCCGCATCTACAAAGCCTTGCTTTGTCTGTGAGGACAAAGTCTCTCAGGCTACCA 398
QY 408 CTATGGGGTCAGCCGCTGTGAGGGCTGCAAGGGCTTTCTCCGCGCAGCATCCAGGAAGAA 467
DB 399 CTATGGGGTCAGCCGCTGTGAGGGCTGCAAGGGCTTTCTCCGCGCAGCATCCAGGAAGAA 458
QY 468 CATGGTCTACACGTGTACCCGGGACAAAGTGTGCATCATCAACAGGTGACCCGGAACCC 527
DB 459 CATGGTCTACACGTGTACCCGGGACAAAGTGTGCATCATCAACAGGTGACCCGGAACCG 518
QY 528 CTGCGAGTACTCGCGACTGCAAGAGTGTCTTTGAAAGTGGGCGATGTCCAAAGGAGTCTGTGAG 587
DB 519 CTGCGAGTACTCGCGACTGCAAGAGTGTCTTTGAAAGTGGGCGATGTCCAAAGGAGTCTGTGAG 578
QY 588 AAACGACCGAAACAAAGAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTGTGAGAGTACAC 647
DB 579 AAACGACCGAAACAAAGAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTGTGAGAGTACAC 638
QY 648 GCTACACCGCGAGGTGGGGAGTCAATTGAGAAGGTGGCAAGCGCACCGACCGAAGACCTT 707
DB 639 GCTACACCGCGAGGTGGGGAGTCAATTGAGAAGGTGGCAAGCGCACCGACCGAAGACCTT 698
QY 708 CCCTGSCCCTCTGCCAGCTGGGCAATACACTACGAACAAACAGCTCAGAACAACTGTCTC 767
DB 699 CCCTGSCCCTCTGCCAGCTGGGCAATACACTACGAACAAACAGCTCAGAACAACTGTCTC 758
QY 768 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAATCTTCCACCAAGTGCATTAAGAC 827
DB 759 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAATCTTCCACCAAGTGCATTAAGAC 818
QY 828 TGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAACACCTCACCACCTCAGGACGATACAC 887
DB 819 TGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAACACCTCACCACCTCAGGACGATACAC 878
QY 888 CCTCCTCAAGGCTGCCTCGCTGGACATCTCTGATCTCGGGATCTGCAGCGGTACACGCC 947
DB 879 CCTCCTCAAGGCTGCCTCGCTGGACATCTCTGATCTCGGGATCTGCAGCGGTACACGCC 938
QY 948 CGAGCAGACACCATGACCTTCTCGGACGGGCTGACCCCTGAACCGGACCCAGATGACAA 1007
DB 939 CGAGCAGACACCATGACCTTCTCGGACGGGCTGACCCCTGAACCGGACCCAGATGACAA 998
QY 1008 CGCTGGCTTGGGCCCTTCACCGACCTGGTCTTTGGCTTCGCCCAACCAAGCTGTGCCCT 1067
DB 999 CGCTGGCTTGGGCCCTTCACCGACCTGGTCTTTGGCTTCGCCCAACCAAGCTGTGCCCT 1058
QY 1068 GGAGATGGATGATCGGAGAGCGGGCTGCTCAGCGCCATCTGCCTCATCTCTCGGAGACCG 1127
DB 1059 GGAGATGGATGATCGGAGAGCGGGCTGCTCAGCGCCATCTGCCTCATCTCTCGGAGACCG 1118
QY 1128 CCAGGACCTGGAGCAGCGGACCGGGGTGGACATGCTGCAGGAGCGGCTGTCTGGAGGCGCT 1187
DB 1119 CCAGGACCTGGAGCAGCGGACCGGGGTGGACATGCTGCAGGAGCGGCTGTCTGGAGGCGCT 1178

QY 1188 AAAGTCTACGTGCGGAAGCGGAGGCCAGCCGCCACATGTTCCCAAGATGCTAAT 1247
DB 1179 AAAGTCTACGTGCGGAAGCGGAGGCCAGCCGCCACATGTTCCCAAGATGCTAAT 1238
QY 1248 GAAGATTACTGACCTGCGAAGCATCAGCCAGGGGGTGTAGCGGTGTATCAGCTGAA 1307
DB 1239 GAAGATTACTGACCTGCGAAGCATCAGCCAGGGGGTGTAGCGGTGTATCAGCTGAA 1298
QY 1308 GATGAGATCCCGGGCTCCATGCCGCTCTCATCCAGGAAATTTGGAGAACTCAGAGGG 1367
DB 1299 GATGAGATCCCGGGCTCCATGCCGCTCTCATCCAGGAAATTTGGAGAACTCAGAGGG 1358
QY 1368 COTGGACACTGTAGCGGACACCGGGGGTGGGGGGCGGGGGTGGCGCTGCCCC 1427
DB 1359 COTGGACACTGTAGCGGACACCGGGGGTGGGGGGCGGGGGTGGCGCTGCCCC 1418
QY 1428 CCGCGCAGCAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCAACAGAGAGCAGCCCGGCCAC 1487
DB 1419 CCGCGCAGCAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCAACAGAGAGCAGCCCGGCCAC 1478
QY 1488 CCACCTCCCGGTGACCGCCACCGCAGCATGACACAGCCCTCGCCCTCCCGCCCGGCTTTT 1547
DB 1479 CCACCTCCCGGTGACCGCCACCGCAGCATGACACAGCCCTCGCCCTCCCGCCCGGCTTTT 1538
QY 1548 CTCTGCTTTTACCGACCATGTGACCCCGCAGCCCTGCCCCCAGCTGCCCCCTCCCGG 1607
DB 1539 CTCTGCTTTTACCGACCATGTGACCCCGCAGCCCTGCCCCCAGCTGCCCCCTCCCGG 1598
QY 1608 GCAGTACTGGGACCTTTCCTGGGGACCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667
DB 1599 GCAGTACTGGGACCTTTCCTGGGGACCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658
QY 1668 AGGCTTGGGCCCTCAGTGGACTGCTGCTCCACAGCCCTGGGTGAGTGTGAGGAGGAGGAG 1727
DB 1659 AGGCTTGGGCCCTCAGTGGACTGCTGCTCCACAGCCCTGGGTGAGTGTGAGGAGGAGGAG 1718
QY 1728 GCAGGAACTGAGTGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1787
DB 1719 GCAGGAACTGAGTGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1778
QY 1788 TTTCATCAAGACACCCCTCTGCCCCAGCTCACCACATCTTCATCACCAGCAACCCAGGAC 1847
DB 1779 TTTCATCAAGACACCCCTCTGCCCCAGCTCACCACATCTTCATCACCAGCAACCCAGGAC 1838
QY 1848 TTGGCTCCCGCATCTCAGAACTCAGAGCCATGTCTCCCGAGCTGGGAGACCTCAACCT 1907
DB 1839 TTGGCTCCCGCATCTCAGAACTCAGAGCCATGTCTCCCGAGCTGGGAGACCTCAACCT 1898
QY 1908 CCCCCCTGCTCGGTGGTGACAGAGGGGTGGGACAGGGGGGGGGGGTTCGCCCTGTAC 1967
DB 1899 CCCCCCTGCTCGGTGGTGACAGAGGGGTGGGACAGGGGGGGGGGGTTCGCCCTGTAC 1958
QY 1968 ATACCTGCTCATACCAACCCAGGATTAATTTCTCCTGCTGTTTGTGTTTATTTTAAATTT 2027
DB 1959 ATACCTGCTCATACCAACCCAGGATTAATTTCTCCTGCTGTTTGTGTTTATTTTAAATTT 2018
QY 2028 TTTTCTTTTGTATTTTAAATAAGAAATTTTCAATTTTAAAGCAAAAAAATAAAAAA 2085
DB 2019 TTTTCTTTTGTATTTTAAATAAGAAATTTTCAATTTTAAAGCAAAAAAATAAAAAA 2076

RESULT 11
109348
LOCUS I09348 2940 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 1 from Patent WO 8905355.
ACCESSION I09348
VERSION I09348.1 GI:587943
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2940)

AUTHORS Evans, R.M., Giguere, V., Ong, E.S., Segui, P.S., Umesono, K. and Thompson, C.C.
JOURNAL Patent: WO 8905355-A 1 15-JUN-1989;
FEATURES Location/Qualifiers
source 1. 2940
BASE COUNT 594 a 1013 c 787 g 546 t
ORIGIN

Query Match 85.28; Score 1777.2; DB 6; Length 2940;
Best Local Similarity 99.3%; Pred. No. 68-312;
Matches 1785; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 288 CTCATTGAGACCCAGACAGCAGTTCGAGAGATGAGTGGCCAGCCCTCCCTCGCCACC 347
DB 279 CACCATGAGACCCAGACAGCAGTTCGAGAGATGAGTGGCCAGCCCTCCCTCGCCACC 338
QY 348 CCCTCTACCCCGCATCTCAAGCCCTTGTCTGTCTCAGGACAAAGTCTCAGGTTACCA 407
DB 339 CCCTCTACCCCGCATCTCAAGCCCTTGTCTGTCTCAGGACAAAGTCTCAGGTTACCA 398
QY 408 CTATGGGTGAGCGCTGTGAGGGTGAAGGGCTTCTTCGCCCGCAGCATCCAGAAGAA 467
DB 399 CTATGGGTGAGCGCTGTGAGGGTGAAGGGCTTCTTCGCCCGCAGCATCCAGAAGAA 458
QY 468 CATGGTGTACAGCTGTACCCGGGACAAAGTGCATCATCAACAAGGTGACCCGGACCC 527
DB 459 CATGGTGTACAGCTGTACCCGGGACAAAGTGCATCATCAACAAGGTGACCCGGACCC 518
QY 528 CTGCGAGTACTGCGGACTGAGAGTGTCTTGAAGTGGGCAAGTGTCCAGAGTCTGTGAG 587
DB 519 CTGCGAGTACTGCGGACTGAGAGTGTCTTGAAGTGGGCAAGTGTCCAGAGTCTGTGAG 578
QY 588 AAACGACCGAAGCAAGAGAGAGTGGCCCAAGCCCGAGTCTCTGAGAGCTATAC 647
DB 579 AAACGACCGAAGCAAGAGAGAGTGGCCCAAGCCCGAGTCTCTGAGAGCTATAC 638
QY 648 GCTGACCGGAGGTGGGGAGCTCATTTGAGAGTGGGCAAGGTCACCAAGCAACCTT 707
DB 639 GCTGACCGGAGGTGGGGAGCTCATTTGAGAGTGGGCAAGGTCACCAAGCAACCTT 698
QY 708 CCCTGCGGCTCTGCGAGTGGGCAATACACTACGAAACAGCTCAGAACACAGTGTCTC 767
DB 699 CCCTGCGGCTCTGCGAGTGGGCAATACACTACGAAACAGCTCAGAACACAGTGTCTC 758
QY 768 TCTGACATTTGACCTCTGGGCAAGTTCAGTGAAGTCTCCACCAAGTGCATATTAAGAC 827
DB 759 TCTGACATTTGACCTCTGGGCAAGTTCAGTGAAGTCTCCACCAAGTGCATATTAAGAC 818
QY 828 TGTGAGTTGCCAAGAGCTGCCGGCTTACACACCTCACCATCGCGGACCATAC 887
DB 819 TGTGAGTTGCCAAGAGCTGCCGGCTTACACACCTCACCATCGCGGACCATAC 878
QY 888 CCTCTCAAGCTGCTGCTGGAGATCTCTGATCTCTGGGATCTGCACGCGTACAGGC 947
DB 879 CCTCTCAAGCTGCTGCTGGAGATCTCTGATCTCTGGGATCTGCACGCGTACAGGC 938
QY 948 CGAGCAGGACACCATGACCTTCTCGGAGCGGTGACCTTGAACCGGACCCAGATGCACAA 1007
DB 939 CGAGCAGGACACCATGACCTTCTCGGAGCGGTGACCTTGAACCGGACCCAGATGCACAA 998
QY 1008 CGCTGGCTTCGGCCCTTACCGACCTGGTCTTTGCTTCGCCATCGCAACAGCTGCTGCCCT 1067
DB 999 CGCTGGCTTCGGCCCTTACCGACCTGGTCTTTGCTTCGCCATCGCAACAGCTGCTGCCCT 1058
QY 1068 GGAGATGATATCGGAGAGCGGGCTGCTCAGCGGCTCTGCTCTGCTGCGGAGACCG 1127
DB 1059 GGAGATGATATCGGAGAGCGGGCTGCTCAGCGGCTCTGCTCTGCTGCGGAGACCG 1118
QY 1128 CCAGGACCTGGAGACCGCGGCTGGAGATGCTGAGGAGCGGCTGCTGAGGCGCT 1187
DB 1119 CCAGGACCTGGAGACCGCGGCTGGAGATGCTGAGGAGCGGCTGCTGAGGCGCT 1178

QY 1188 AAAGTCTACTGTCGCGAAGCGAGGCGCCACCATGTTCCCAAGATGCTTAAT 1247
DB 1179 AAAGTCTACTGTCGCGAAGCGAGGCGCCACCATGTTCCCAAGATGCTTAAT 1238
QY 1248 GAAGATTACTGACCTGCGAAGCATCAGCGCCAAAGGGGCTGAGGGGTGATCAGCTGAA 1307
DB 1239 GAAGATTACTGACCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGTGATCAGCTGAA 1298
QY 1308 GATGGAGATCCCGGCTCCATGCCGCTCTCATCCAGGAATGTTGAGAACTCAGAGGG 1367
DB 1299 GATGGAGATCCCGGCTCCATGCCGCTCTCATCCAGGAATGTTGAGAACTCAGAGGG 1358
QY 1368 CTGGAACACTCTGAGCGGACAGCGCGGGGTGGGGGCGGAGCGGGGTGCTGCCCCC 1427
DB 1359 CTGGAACACTCTGAGCGGACAGCGCGGGGTGGGGGCGGAGCGGGGTGCTGCCCCC 1418
QY 1428 CCCGCCAGCAGCTGTAGCCCCAGCTCAGCCCCAGTCCACAGAAAGAGCCGGCCAC 1487
DB 1419 CCCGCCAGCAGCTGTAGCCCCAGCTCAGCCCCAGTCCACAGAAAGAGAGCCGGCCAC 1478
QY 1488 CCACCTCCCGTGCAGCGCCACGCGCCACATGGACAGACGCTCGCCCTCGCCCGGCTTTT 1547
DB 1479 CCACCTCCCGTGCAGCGCCACGCGCCACATGGACAGACGCTCGCCCTCGCCCGGCTTTT 1538
QY 1548 CTCTGCTTTTCTACCGACCATGTACCCGACAGCGCTGCCCGCCACCTGCCCTCCCGG 1607
DB 1539 CTCTGCTTTTCTACCGACCATGTACCCGACAGCGCTGCCCGCCACCTGCCCTCCCGG 1598
QY 1608 GCGACTACTGGGACCTTCCCTGGGGACGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667
DB 1599 GCGACTACTGGGACCTTCCCTGGGGACGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658
QY 1668 AGGCTGGGCGCTCAGTGGAGCTGCTCCACAGCGCTGGGCTGAGCTCAGAGGCGCGAG 1727
DB 1659 AGGCTGGGCGCTCAGTGGAGCTGCTCCACAGCGCTGGGCTGAGCTCAGAGGCGCGAG 1718
QY 1728 GCGAGAACTGAGTGGGCGCTTGGGTCTCAGAGTGGGTCTCGGGGCTCGTG 1787
DB 1719 GCGAGAACTGAGTGGGCGCTTGGGTCTCAGAGTGGGTCTCGGGGCTCGTG 1778
QY 1788 TTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCACCAAGCCAGGAC 1847
DB 1779 TTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCACCAAGCCAGGAC 1838
QY 1848 TTGGCTCCCGCTCAGTCAAGACTCAGAGCCATGCTCCCGAGCTGGGGAGGAGGAGGAG 1907
DB 1839 TTGGCTCCCGCTCAGTCAAGACTCAGAGCCATGCTCCCGAGCTGGGGAGGAGGAGGAG 1898
QY 1908 CCCCTGCTCGTGGTGGTGGAGAGGGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1967
DB 1899 CCCCTGCTCGTGGTGGTGGAGAGGGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1958
QY 1968 ATACCTGCTCAGTACCAACCCAGGATTAATCTCGTGGTGGTGGTGGTGGTGGTGGTGGT 2037
DB 1959 ATACCTGCTCAGTACCAACCCAGGATTAATCTCGTGGTGGTGGTGGTGGTGGTGGTGGT 2018
QY 2028 TTTTGGTTGGATTTTAAAGAAATTTTCATTTTAAGCACAATAAAAAA 2085
DB 2019 TTTTGGTTGGATTTTAAAGAAATTTTCATTTTAAGCACAATAAAAAA 2076

RESULT 12
AK098172

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

AK098172 Homo sapiens cDNA FLJ40853 fis, clone TRACH2015486, highly similar to RETINOIC ACID RECEPTOR ALPHA.
AK098172 AK098172.1 GI:21758129
oligo capping; fis (full insert sequence).
Homo sapiens trachea cDNA to mRNA, clone_lib:TRACH2
clone:TRACH2015486.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```
|||||
Db 1714 TGGGGGCTCGTGTTCATCAGACACCCCTCGCCAGCTCACCACATCTTCATCACCAG 1773
Qy 1835 CAACGCCAGGACTTGGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGG 1894
Db 1774 CAACGCCAGGACTTGGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGG 1833
Qy 1895 GGAACCTCAACCTCCCGCTCGCTGGTGTGACAGAGGGGTGGGACAGGGGGGGGG 1954
Db 1834 GGAACCTCAACCTCCCGCTCGCTGGTGTGACAGAGGGGTGGGACAGGGGGGGGG 1893
Qy 1955 GTTCCCTCTGACATACCTCGCATACCAACCCAGGATTAATTCGCTGGTTTGT 2014
Db 1894 GTTCCCTCTGACATACCTCGCATACCAACCCAGGATTAATTCGCTGGTTTGT 1953
Qy 2015 TTTATTTAAATTTTGTGTTGATTTTAAATTAAGAAATTTTCAATTTAAGCAC 2069
Db 1954 TTTATTTAAATTTTGTGTTGATTTTAAATTAAGAAATTTTCAATTTAAGCAC 2008

RESULT 13
HSU41742 2073 bp mRNA linear PRI 17-MAY-1996
LOCUS
DEFINITION
NPW-RAR long form mRNA, complete cds.
ACCESSION
D41742 GI:1314307
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2073)
Redner, R.L., Rush, E.A., Faas, S., Rudert, W.A. and Corey, S.J.
The t(5;17) variant of acute promyelocytic leukemia expresses a
nucleophosmin-retinoic acid receptor fusion
Blood 87 (3), 882-886 (1996)
JOURNAL
MEDLINE
PUBMED
REFERENCE
8562957
2 (bases 1 to 2073)
Redner, R.L.
Direct Submision
Submitted (30-NOV-1995) Robert L. Redner, Medicine, University of
Pittsburgh, E1058 BST, 211 Lothrop Street, Pittsburgh, PA 15213,
USA
FEATURES
Location/Qualifiers
1. 2073
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="acute promyelocytic leukemia cells with
t(5;17)(q32;q11) translocation"
/tissue_type="bone marrow"
84..1775
/codon_start=1
/product="nucleophosmin-retinoic acid receptor alpha
fusion protein NPW-RAR long form"
/protein_id="AAB00112.1"
/db_xref="GI:1314308"
/translaton="MEDSMDMSPRLPQNYLFGCEKADKDYHPKVDNENEHOLS
RTVSLGAGAKDELHIVERAANYBESPIKVTLATLKMVSQVPTSLGPEITPPVYLR
KCGSPVHISQHLVAVEDAESEDEEDENRVLPMKPKRGLKHMFGDLVCSKRLA
AIEQSSSEIEVPSPDPLRIYKPCFVCQDKSSGYHYGVSAECGCKGFRRSIQ
KNMYTTRHRDKNCRKYNRQYCRQKGFYVGMKSSEYRNDNRKXKKEVPKPCS
ESYTLTPVGELEIKRKAHOETFPALCOLGKFTNNSEQRVSLDIDLWKFSELS
KCIKTVEFAKQLGFTTLTADQITLLKACLDLILIRICTRYTPEDDTWPSGLT
LNQTMHNAEGPLDVLVAFANQLLPLEMDAETGLISALICLGQRDLQEPQPRVD
MLQEPLEALKYVYVRKRPSPRPFPMKIMKIDLRSISAKGAERVITLKMIEGMP
PLIQEMLENSGLDTLSQPGGGGGGLAPPGGCSPSLSPSSNRSSPATHSP"
BASE COUNT 465 a 612 c 605 g 391 t
ORIGIN
```

Query Match 71.9%; Score 1500.8; DB 9; Length 2073;
Best Local Similarity 98.3%; Pred. No. 7.le-262;

```
Matches 1517; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Qy 255 GCTCCGGACTCCGCTTTGGAATGGCTCAAAACCACTCCATTGAGACCCAGAGCAGCTTC 314
Db 530 GTTTCGGGATTTGGTGTGAGTTGGCAATTTGGCAGCCATTGAGACCCAGAGCAGCTTC 589
Qy 315 TCAAGAGATAGTGGCCAGCCCTCCCTCGCCACCCCTCTACCCCGGATCTACAAAGCCCTG 374
Db 590 TGAAGAGATAGTGGCCAGCCCTCCCTCGCCACCCCTCTACCCCGGATCTACAAAGCCCTG 649
Qy 375 CTTTGTCTCTCAGGACAAGTCTCAGGCTACCACTATGAGGGTCAAGGCTGTGAGGCTG 434
Db 650 CTTTGTCTCTCAGGACAAGTCTCAGGCTACCACTATGAGGGTCAAGGCTGTGAGGCTG 709
Qy 435 CAAGGCTCTCTCCCGCGCAGCATCCAGAGAAACATGGTGTACACGTGTACCCGGGACAA 494
Db 710 CAAGGCTCTCTCCCGCGCAGCATCCAGAGAAACATGGTGTACACGTGTACCCGGGACAA 769
Qy 495 GAACTGCATCATCAACAAGGTGACCCGGGAACCCCTGCCAGTACTGCCGACTGCAGAAAGT 554
Db 770 GAACTGCATCATCAACAAGGTGACCCGGGAACCCCTGCCAGTACTGCCGACTGCAGAAAGT 829
Qy 555 CTTTGAAGTGGGATCTCCAAAGAGTCTGTGAGAAACGACCCGAAACAAAGAAAGAAAGGA 614
Db 830 CTTTGAAGTGGGATCTCCAAAGAGTCTGTGAGAAACGACCCGAAACAAAGAAAGAAAGGA 889
Qy 615 GGTGCCAAAGCCGAGTGTCTGAGAGCTACACGCTGACCCGGAGGTGGGGGAGCTCAT 674
Db 890 GGTGCCAAAGCCGAGTGTCTGAGAGCTACACGCTGACCCGGAGGTGGGGGAGCTCAT 949
Qy 675 TGAGAGGTGGCCAAAGCCGACCCAGGAACCTTCCTCCCTGCCCTCTGCCAGCTGGGCAATA 734
Db 950 TGAGAGGTGGCCAAAGCCGACCCAGGAACCTTCCTCCCTGCCCTCTGCCAGCTGGGCAATA 1009
Qy 735 CACTACGAACAACAGCTCAGAACAAAGTGTCTCTCTGGACATTTGACCTCTGGGCAAGTT 794
Db 1010 CACTACGAACAACAGCTCAGAACAAAGTGTCTCTCTGGACATTTGACCTCTGGGCAAGTT 1069
Qy 795 CAGTGAACCTCTCCCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGG 854
Db 1070 CAGTGAACCTCTCCCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGG 1129
Qy 855 CTTACACCCCTCACCATGCCGACCAAGATCACCTCTCTCAAGGCTCCCTGCTGGACAT 914
Db 1130 CTTACACCCCTCACCATGCCGACCAAGATCACCTCTCTCAAGGCTCCCTGCTGGACAT 1189
Qy 915 CTTGATCTCTCGGATCTGCACGCGGTACACGCCCGAGCAGGACACCATGACCTTCTCGGA 974
Db 1190 CTTGATCTCTCGGATCTGCACGCGGTACACGCCCGAGCAGGACACCATGACCTTCTCGGA 1249
Qy 975 CGGGCTGACCCCTGAACCGGACCCAGATGCACACGCTGGCTTGGGCCCTTCACCGACCT 1034
Db 1250 CGGGCTGACCCCTGAACCGGACCCAGATGCACACGCTGGCTTGGGCCCTTCACCGACCT 1309
Qy 1035 GGTCTTTGCTTCCCAACCAAGCTGTGCTCCCTGGAGATGGATGATCGGAGACGGGGCT 1094
Db 1310 GGTCTTTGCTTCCCAACCAAGCTGTGCTCCCTGGAGATGGATGATCGGAGACGGGGCT 1369
Qy 1095 GCTCAGCGCCATCTGCTCATCTGCGGAGACCCGACGAGCTGGAGCAGCCGCGGACCGGGT 1154
Db 1370 GCTCAGCGCCATCTGCTCATCTGCGGAGACCCGACGAGCTGGAGCAGCCGCGGACCGGGT 1429
Qy 1155 GGACATGCTGACAGAGCCGCTGTGAGGCGCTTAAAGTCTACGTGGGAAGCGGAGGCC 1214
Db 1430 GGACATGCTGACAGAGCCGCTGTGAGGCGCTTAAAGTCTACGTGGGAAGCGGAGGCC 1489
Qy 1215 CAGCGCCGCCCATGTTTCCCAAGATGCTTAATGAAGTACTGACCTGCCAAGCATCAG 1274
Db 1490 CAGCGCCGCCCATGTTTCCCAAGATGCTTAATGAAGTACTGACCTGCCAAGCATCAG 1549
Qy 1275 CGCCAAAGGGGCTGAGCGGGTGTACGCTGAAGATGGAGATCCCGGGCTCCATGCCGCC 1334
Db 1550 CGCCAAAGGGGCTGAGCGGGTGTACGCTGAAGATGGAGATCCCGGGCTCCATGCCGCC 1609
```

```
QY 1335 TCTCATCCAGGAATGTTGGAGAACTCAGAGGGCTTGACACTCTGACGCGACAGCCGGG 1394
Db 1610 TCTCATCCAGGAATGTTGGAGAACTCAGAGGGCTTGACACTCTGACGCGACAGCCGGG 1669
QY 1395 GGTGGGGGGGGACGGGGGTGCGCTGCCCCCCCCCGCCAGGAGCTGTAGCCCGAGCCT 1454
Db 1670 GGTGGGGGGGGACGGGGGTGCGCTGCCCCCCCCCGCCAGGAGCTGTAGCCCGAGCCT 1729
QY 1455 CAGCCCCAGCTCCAAACAGAGCAGCCGGCCACCACTCCCGGTGACCGCCAGCCACACA 1514
Db 1730 CAGCCCCAGCTCCAAACAGAGCAGCCGGCCACCACTCCCGGTGACCGCCAGCCACACA 1789
QY 1515 TGGACACAGCCCTCGCCCTCGCCCCCGGCTTTCTCTGCTTTTCTACCGACCATGTGACC 1574
Db 1790 TGGACACAGCCCTCGCCCTCGCCCCCGGCTTTCTCTGCTTTTCTACCGACCATGTGACC 1849
QY 1575 CCGACACAGCCCTCGCCCTCGCCCCCGGCTTTCTCTGCTTTTCTACCGACCATGTGACC 1634
Db 1850 CCGACACAGCCCTCGCCCTCGCCCCCGGCTTTCTCTGCTTTTCTACCGACCATGTGACC 1909
QY 1635 CGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1694
Db 1910 CGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1969
QY 1695 CTCCACAGCCCTGGGCTGACGCTCAGAGCCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAG 1754
Db 1970 CTCCACAGCCCTGGGCTGACGCTCAGAGCCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAG 2029
QY 1755 CTGGGCTCAGGATGGGTCTGGGGGCTCTGTGTATCAAGAC 1798
Db 2030 CTGGGCTCAGGATGGGTCTGGGGGCTCTGTGTATCAAGAC 2073

RESULT 14
HSRAR HSRAR 1920 bp mRNA linear PRI 12-SEP-1993
LOCUS Human mRNA for retinoic acid receptor.
DEFINITION X06538
ACCESSION X06538
VERSION X06538.1 GI:35873
KEYWORDS DNA binding protein; hormone receptor; receptor; retinoic acid receptor.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 260 to 1920)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Pekovich, M., Brand, N.J., Krust, A. and Chambon, P.
A human retinoic acid receptor which belongs to the family of
nuclear receptors
JOURNAL Nature 330 (6147), 444-450 (1987)
MEDLINE 88063872
PUBMED 2825025
REFERENCE 2 (bases 1 to 420)
AUTHORS Chambon, P.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1988)
COMMENT cell line-MCF-7; library-lambda gt10; clone-p63.
FEATURES
source
1. 1920
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
317..1615
/note="retinoic acid receptor (AA 1 - 432)"
/codon_start=1
/protein_id="CAA29787.1"
/db_xref="GI:35874"
/db_xref="SWISS-PROT:P10276"
/translation="MLGSLSPGALTLQHLQVPSGYSTPSPATIEQSSSEIIVPS
PPSPPLRIYKPCFVQDKSSGIHYGSACEGKGFFRRSIQKNMYVTCRRKNCII
NKYTRNRCQYRLQKQFVGMGSKESYVNRDNKKKKEVPKPECSSEYTLTPEVGELEIK
VRKHAQETFPALQKQYITNNSEQRSLDIDLWPKFSELSTKCIKTVEFAKQLPG
FTTLTIADQITLLKAACLDILILRICTRYTPEQDTWTFTSDGLTLNRTMHNAGFPLT
```

```
DLVAFANQLLPLEMDAETGLLSAICLICDRODLEQPDVRDMLQELLEALKVYVR
KRPSRPHMFPKMLMKITDLRSISAKGAERVITLKMETPSGMPPLIOEMLENSEGLDT
LSGPGGGGGGGGLAPPGGSPSLSPSSNRSSPATHSP"
488..685
/note="put. DNA-binding domain"
763..764
/note="gc was cg in [1]"
/citation=[1]
824..1483
/note="put. ligand-binding domain"
1395
/note="g was c in [1]"
/citation=[1]
1733
/note="g was c in [1]"
/citation=[1]
BASE COUNT 386 a 638 c 566 g 330 t
ORIGIN

Query Match 71.9%; Score 1499.8; DB 9; Length 1920;
Best Local Similarity 99.5%; Pred. No. 1.1e-261;
Matches 1504; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 288 CTCCATTGAGACCCAGACGAGCAGTTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC 347
Db 403 CACCATTGAGACCCAGACGAGCAGTTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC 462
QY 348 CCCTTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAGTCTCAGGCTACCA 407
Db 463 CCCTTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAGTCTCAGGCTACCA 522
QY 408 CTATGGGCTACGCGCTGTGAGGGCTGCAAGGGCTTCTTCGCGCCGAGCATCCAGAAGAA 467
Db 523 CTATGGGCTACGCGCTGTGAGGGCTGCAAGGGCTTCTTCGCGCCGAGCATCCAGAAGAA 582
QY 468 CATGTTGTACAGTGTACCCGGGACAAAGATGTCATCATCAACAAGGTGACCCGGAACCC 527
Db 593 CATGTTGTACAGTGTACCCGGGACAAAGATGTCATCATCAACAAGGTGACCCGGAACCC 642
QY 528 CTGCCAGTACTGCCAGTGCAGAGTGTCTTGAAGTGGGCATGTCCAGGAGTCTGTGAG 587
Db 643 CTGCCAGTACTGCCAGTGCAGAGTGTCTTGAAGTGGGCATGTCCAGGAGTCTGTGAG 702
QY 588 AAACGACCGAAACAGAAAGAGAGGAGTGCACCAAGCCGAGTGTCTGTGAGAGCTACAC 647
Db 703 AAACGACCGAAACAGAAAGAGAGGAGTGCACCAAGCCGAGTGTCTGTGAGAGCTACAC 762
QY 648 GCTGACCGCGGAGGTGGGGGAGCTCATTTGAGAAGGTGCGCAAGGCGCACAGGAAACCTT 707
Db 763 GCTGACCGCGGAGGTGGGGGAGCTCATTTGAGAAGGTGCGCAAGGCGCACAGGAAACCTT 822
QY 708 CCCTGCCCTCTGCCAGTGGGCAATATACACTACGAAACAGCTCAGAACACAGTGTCTC 767
Db 823 CCCTGCCCTCTGCCAGTGGGCAATATACACTACGAAACAGCTCAGAACACAGTGTCTC 882
QY 768 TCTGACATTCACCTCTCGGACAGTTCAGTGAACCTTCCACCAAGTGCATCATTAAGAC 827
Db 882 TCTGACATTCACCTCTCGGACAGTTCAGTGAACCTTCCACCAAGTGCATCATTAAGAC 942
QY 828 TGTGAGTTCGCCAGAGCTGCCCGGCTTACACCCCTTCCACCATCGCGGACCATCATAC 887
Db 943 TGTGAGTTCGCCAGAGCTGCCCGGCTTACACCCCTTCCACCATCGCGGACCATCATAC 1002
QY 888 CCTGCTCAAGGCTGCTGCTGCGGACATCGTGATCTGGGATCTGACGCGGTACAGCC 947
Db 1003 CCTGCTCAAGGCTGCTGCTGCGGACATCGTGATCTGGGATCTGACGCGGTACAGCC 1062
QY 948 CGAGCAGACACCATGAGCTTCTCGGAGGGCTGACCTGAAACCGGACCCAGATGCACAA 1007
Db 1063 CGAGCAGACACCATGAGCTTCTCGGAGGGCTGACCTGAAACCGGACCCAGATGCACAA 1122
QY 1008 CGCTGGCTTCGGCCGCCCTTCACCGAGCTGTCTTTGCTTTCGCAACACAGCTGCTGCCCT 1067
```


QY 830 TGGAGTTCCGCAAGCAGTGCCCGGCTTCACACCCCTCACCATCGCCGACAGATCACCC 889
Db 976 TGGAGTTCCGCAAGCAGTGCCCGGCTTCACACCCCTCACCATCGCCGACAGATCACCC 1035
QY 890 TCCTCAAGGCTGCTGCTGACATCCTGATCCTGCGGATCTGCAGGGTACAGCCCG 949
Db 1036 TCCTCAAGGCTGCTGCTGACATCCTGATCCTGCGGATCTGCAGGGTACAGCCCG 1095
QY 950 AGCAGGACACCATGACCTTCCTCGGAGGGCTGACCTGAACCGGACCCAGATGCACACG 1009
Db 1096 AGCAGGACACCATGACCTTCCTCGGAGGGCTGACCTGAACCGGACCCAGATGCACACG 1155
QY 1010 CTGCTTCGGCCCTCACCAGACTGCTTTGCTTCGCCAACAGCTGCTGCCCTGG 1069
Db 1156 CTGCTTCGGCCCTCACCAGACTGCTTTGCTTCGCCAACAGCTGCTGCCCTGG 1215
QY 1070 AGATGGATGATGGGAGACGGGCTGCTCAGCGCATCTGCCTCATCTGCGGAGACGGCC 1129
Db 1216 AGATGGATGATGGGAGACGGGCTGCTCAGCGCATCTGCCTCATCTGCGGAGACGGCC 1275
QY 1130 AGGACCTGGAGAGCCGACCGGCTGGACATGCTGCAGGAGCGCTGCTGGAGCGCTAA 1189
Db 1276 AGGACCTGGAGAGCCGACCGGCTGGACATGCTGCAGGAGCGCTGCTGGAGCGCTAA 1335
QY 1190 AGGTCTACGTGCGGAAGCGGAGGCCAGCCGCCACATGTTCCCAAGATGCTAATGA 1249
Db 1336 AGGTCTACGTGCGGAAGCGGAGGCCAGCCGCCACATGTTCCCAAGATGCTAATGA 1395
QY 1250 AGATTACTGACCTGCGAAGCATCAGCCCAAGGGGGTGAAGGGGTGATCAGCTGAAGA 1309
Db 1396 AGATTACTGACCTGCGAAGCATCAGCCCAAGGGGGTGAAGGGGTGATCAGCTGAAGA 1455
QY 1310 TGGAGATCCCGGCTCCATCGCCCTCTCATCCAGGAATGTTGGAGAACTCAGAGGGCC 1369
Db 1456 TGGAGATCCCGGCTCCATCGCCCTCTCATCCAGGAATGTTGGAGAACTCAGAGGGCC 1515
QY 1370 TGGACACTCTGAGCGGACAGCCGGGGTGGGGGGGAGCGGGTGGCTGCCCGCCC 1429
Db 1516 TGGACACTCTGAGCGGACAGCCGGGGTGGGGGGGAGCGGGTGGCTGCCCGCCC 1575
QY 1430 CGCCAGGACAGCTGATGCCCGCAGCTCAGCCCCAGCTCCACAGAGAGCCCGCCACCC 1489
Db 1576 CGCCAGGACAGCTGATGCCCGCAGCTCAGCCCCAGCTCCACAGAGAGCCCGCCACCC 1635
QY 1490 ACTCCCGCTGACCGCCACGCCACATGGACACAGCCCTCGCCCTCGCCCGCCCTTTCT 1549
Db 1636 ACTCCCGCTGACCGCCACGCCACATGGACACAGCCCTCGCCCTCGCCCGCCCTTTCT 1695
QY 1550 CTGCTTTCTACCGACCATGTGACCCCGCAGCCAGCCCTGCCCTCGCCCGCC 1609
Db 1696 CTGCTTTCTACCGACCATGTGACCCCGCAGCCAGCCCTGCCCTCGCCCGCC 1755
QY 1610 AGTACTGGGAGCTTCCCTGGGGAGGGAGGGAGGAGGAGCGACTCCTTGGACAGAG 1669
Db 1756 AGTACTGGGAGCTTCCCTGGGGAGGGAGGGAGGAGGAGGAGGAGCGACTCCTTGGACAGAG 1815
QY 1670 GCCTGGGCGCTCAGTGGACTGCTGCTCCACAGCTGGCTGACCTCAGAGCGCCAGGC 1729
Db 1816 GCCTGGGCGCTCAGTGGACTGCTGCTCCACAGCTGGCTGACCTCAGAGCGCCAGGC 1875
QY 1730 CAGGAACAGTGAAGCCCTGCTGCTGCTGCTCAGGATGGGTCTCAGGATGGGGCTCGTGT 1789
Db 1876 CAGGAACAGTGAAGCCCTGCTGCTGCTGCTCAGGATGGGTCTCAGGATGGGGCTCGTGT 1935
QY 1790 CATCAAGAC 1798
Db 1936 CATCAAGAC 1944

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2003, 05:30:18 ; Search time 62 seconds

(without alignments)
1518.767 Million cell updates/sec

Title: US-09-691-220-2

Perfect score: 2437

Sequence: 1 MYSEVEGGTPNPFVLVDF.....SCSPSLSPSSNRSSPATHSP 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp Vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2349	96.4	459	11 Q9QWJ1	Q9QWJ1 rattus norv
2	2112	86.7	797	4 Q15156	Q15156 homo sapien
3	2092	85.8	462	11 P97513	P97513 mus spretus
4	2072.5	85.0	430	11 Q8VHB8	Q8VHB8 mesocricetu
5	2026.5	83.2	448	13 Q92019	Q92019 xenopus lae
6	2021	82.9	457	13 Q90272	Q90272 brachydanio
7	2008.5	82.4	444	13 Q91391	Q91391 brachydanio
8	1998.5	82.0	444	13 Q90271	Q90271 brachydanio
9	1770	72.6	448	11 Q8VHB7	Q8VHB7 mesocricetu
10	1760	72.2	448	13 Q910C7	Q910C7 coturnix co
11	1651	67.7	458	11 Q91YX2	Q91YX2 mus musculu
12	1651	67.7	458	11 Q91VK5	Q91VK5 mus musculu
13	1646.5	67.6	443	11 Q8VHB6	Q8VHB6 mesocricetu
14	1641	67.3	441	13 Q918T3	Q918T3 ambystoma m
15	1628.5	66.8	446	13 Q918T2	Q918T2 ambystoma m
16	1620	66.5	307	11 Q9QWT3	Q9QWT3 rattus norv

17	1613	66.2	426	13 Q90968	Q90968 gallus gall
18	1595.5	65.5	381	4 Q9UJ38	Q9UJ38 homo sapien
19	1474.5	60.5	367	13 Q90273	Q90273 brachydanio
20	1387	56.9	302	11 Q9QWTA	Q9QWTA rattus norv
21	1358	55.7	285	13 Q90Y68	Q90Y68 paralichthy
22	1335	54.8	285	13 Q90Y67	Q90Y67 paralichthy
23	1290	52.9	279	13 Q9DGY6	Q9DGY6 salmo salar
24	1287.5	52.8	286	13 Q90Y70	Q90Y70 paralichthy
25	1234	50.6	287	13 Q90Y69	Q90Y69 paralichthy
26	1173	48.1	434	5 P91780	P91780 polyandroca
27	1151	47.2	291	11 Q54746	Q54746 rattus norv
28	988	40.5	347	5 P91779	P91779 polyandroca
29	950	39.0	224	6 P79300	P79300 sus scrofa
30	683	28.0	130	6 P79298	P79298 sus scrofa
31	652	26.8	410	4 Q96H73	Q96H73 homo sapien
32	627	25.7	141	6 P79297	P79297 sus scrofa
33	625	25.6	387	13 Q90Y21	Q90Y21 salmo salar
34	625	25.6	555	15 Q96594	Q96594 avian eryth
35	625	25.6	582	15 Q96593	Q96593 avian eryth
36	623.5	25.6	396	13 Q90Y22	Q90Y22 salmo salar
37	623.5	25.6	455	15 Q85511	Q85511 avian eryth
38	617	25.3	555	15 Q36200	Q36200 avian eryth
39	612.5	25.1	962	15 Q64895	Q64895 avian eryth
40	562.5	23.1	321	13 Q90Y30	Q90Y30 fugu rubrip
41	553.5	22.7	445	11 Q91X41	Q91X41 mus musculu
42	537	22.0	412	13 Q91840	Q91840 xenopus lae
43	536.5	22.0	545	5 Q9U4L1	Q9U4L1 aedes aegyp
44	534	21.9	149	13 Q918T4	Q918T4 ambystoma m
45	532	21.8	329	13 Q90Y19	Q90Y19 scophthalmu

ALIGNMENTS

RESULT 1

Q9QWJ1 PRELIMINARY; PRT; 459 AA.
ID Q9QWJ1
AC Q9QWJ1
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DE 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Retinoic acid receptor alpha 2 isoform.
GN RAR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-TESTIS;
RX MEDLINE=96301334; PubMed=8722633;
RA Akmal K.M., Dufour J.M., Kim K.H.;
RT "Region-specific localization of retinoic acid receptor-alpha
expression in the rat epididymis";
RL Biol. Reprod. 54:1111-1119(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-TESTIS;
RX MEDLINE=97167633; PubMed=9116160;
RA Akmal K.M., Dufour J.M., Kim K.H.;
RT "Retinoic acid receptor alpha gene expression in the rat testis:
potential role during the prophase of meiosis and in the transition
from round to elongating spermatids";
RL Biol. Reprod. 56:549-556(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-TESTIS;
RX MEDLINE=98151023; PubMed=9492059;
RA Akmal K.M., Dufour J.M., Vo M., Higginson S., Kim K.H.;
RT "Ligand-dependent regulation of retinoic acid receptor alpha in rat
testis: in vivo response to depletion and repletion of vitamin A";
RL Endocrinology 139:1239-1248(1998).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; D15211; AAC23439.1; -.
DR HSSP; P10826; 1HRA.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrmn_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 459 AA; 50928 MW; F03368934C1E3C86 CRC64;

Query Match 96.4%; Score 2349; DB 11; Length 459;
Best Local Similarity 96.9%; Pred. No. 2.7e-195;
Matches 445; Conservative 4; Mismatches 8; Indels 2; Gaps 1;

QY 1 MYSEVEVG--PTPNFVYDFNQNRACLLPEKGLPAGPYSTPLRTPLWNGSNHSIET 58
DB 1 MYSEVEVGLTPAPNPLVYDFNQNRACLLPEKGLPAGPYSTPLRTPLWNGSNHSIET 60

QY 59 QSSSSEIIVSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKNVYT 118
DB 61 QSSSSEIIVSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKNVYT 120

QY 119 CHRDKNCIINKVTRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCSESYTLTPE 178
DB 121 CHRDKNCIINKVTRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCSESYTLTPE 180

QY 179 VGLIEKVRKAHOETFPALCQKGYTNNSEQRVSLDIDLWKFSELSTKCIITVEFA 238
DB 181 VGLIEKVRKAHOETFPALCQKGYTNNSEQRVSLDIDLWKFSELSTKCIITVEFA 240

QY 239 KQLPGFTTITADQITLLKAAACLDILILTRICTRYTPEDQTMFTSDGLTLNRTQMHNAGF 298
DB 241 KQLPGFTTITADQITLLKAAACLDILILTRICTRYTPEDQTMFTSDGLTLNRTQMHNAGF 300

QY 299 PLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRQDLQPPDRVDMQLQEPLEALKVYV 358
DB 301 PLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRQDLQPPDRVDMQLQEPLEALKVYV 360

QY 359 RRRRPSRPHMFKMLMKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSEGLDTL 418
DB 361 RRRRPSQPHMFKMLMKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSEGLDTL 420

QY 419 SGQPGGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457
DB 421 SGQSGGGTRDGGGLAPPGSCPSLSPSSNRSSPATHSP 459

RESULT 2
Q15156 PRELIMINARY; PRT; 797 AA.
AC Q15156;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE PML-RAR protein.
GN PML-RAR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91347368; PubMed-16523368;
RA Kakizuka A., Miller W.H. Jr., Umesono K., Warrell R.P. Jr.,
RA Frankel S.R., Murty V.V., Dmitrovsky E., Evans R.M.;

Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RAR alpha with a novel putative transcription factor, PML;
Cell 66:663-674(1991).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; M73779; AAA60126.1; -.
DR HSSP; P29590; 1BOR.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrmn_receptor.
DR InterPro; IPR000315; Znf_Box.
DR InterPro; IPR001628; Znf_C4steroid.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00643; zf-B_box; 2.
DR Pfam; PF00097; zf-C3HC4; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00336; BBOX; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 797 AA; 89281 MW; 8C16B6E85CDBD214 CRC64;

Query Match 86.7%; Score 2112; DB 4; Length 797;
Best Local Similarity 99.3%; Pred. No. 2e-174;
Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 55 SIETQSSSEIIVSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKN 114
DB 395 AIETQSSSEIIVSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKN 454

QY 115 MYTCHRDKNCIINKVTRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCSESYT 174
DB 455 MYTCHRDKNCIINKVTRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCSESYT 514

QY 175 LTPVEGLIEKVRKAHOETFPALCQKGYTNNSEQRVSLDIDLWKFSELSTKCIIT 234
DB 515 LTPVEGLIEKVRKAHOETFPALCQKGYTNNSEQRVSLDIDLWKFSELSTKCIIT 574

QY 235 VEFAPQLPGFTTITADQITLLKAAACLDILILTRICTRYTPEDQTMFTSDGLTLNRTQMHN 294
DB 575 VEFAPQLPGFTTITADQITLLKAAACLDILILTRICTRYTPEDQTMFTSDGLTLNRTQMHN 634

QY 295 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRQDLQPPDRVDMQLQEPLEAL 354
DB 635 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRQDLQPPDRVDMQLQEPLEAL 694

QY 355 KVYVRRRPSRPHMFKMLMKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSE 414
DB 695 KVYVRRRPSRPHMFKMLMKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSE 754

QY 415 LDTLSGQPGGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457
DB 755 LDTLSGQPGGGGRDGGGLAPPGSCPSLSPSSNRSSPATHSP 797

RESULT 3
P97513 PRELIMINARY; PRT; 462 AA.
ID P97513;
AC P97513;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Retinoic acid receptor-alpha.
GN RARA.

DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Retinoic acid receptor.
 GN RARA.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE-93144180; PubMed-1336976;
 RA Sharpe C.R.;
 RT "Two isoforms of retinoic acid receptor alpha expressed during Xenopus
 RT development respond to retinoic acid."
 RL Mech. Dev. 39:81-93(1992).
 RN [2]
 RN SEQUENCE FROM N.A.
 RA Sharpe C.R.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL; X87363; CAA60791.1; -.
 DR HSSP; P10826; IHRA.
 DR InterPro; IPR000536; Hormone_rec_lig.
 DR InterPro; IPR001723; Stdhrmn_receptor.
 DR Pfam; PF00104; hormone_rec; 1.
 DR Pfam; PF00105; zf-C4; 1.
 DR PRINTS; PR00398; STDRHORMONER.
 DR PROSITE; PS00047; STROIDFINGER.
 DR ProDom; PD000035; Znf_C4steroid; 1.
 DR SMART; SM00430; HOL1; 1.
 DR SMART; SM00399; ZNF_C4; 1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 DR DNA-binding; Nuclear protein; Receptor; Transcription regulation;
 KW Zinc-finger.
 SQ SEQUENCE 448 AA; 50456 MW; 6CF94B31BC57054B CRC64;

 Query Match 83.2%; Score 2026.5; DB 13; Length 448;
 Best Local Similarity 84.9%; Pred. No. 2.4e-167;
 Matches 388; Conservative 22; Mismatches 38; Indels 9; Gaps 3;
 QY 1 MYESVEVGGPTNPFLVDFYQNRACLLPEKGLPAPGYPSTPLRPLWNGSNHSIETQS 60
 DB 1 MYENVDV---SPHYHMDYFYNRQCLWPEKRI---NPYGLTGTQHWSSSNHSIETQS 54
 QY 61 SSSEIIVSPSPPPPLPRIYKPCFVCQDKSSGYHYGVSACEGCKGFFRRIQKNMVTCH 120
 DB 55 TSSEIIVSPSPPPPLPRIYKPCFVCQDKSSGYHYGVSACEGCKGFFRRIQKNMVTCH 114
 QY 121 RDKNCLINKVTRNFCQYCRLOKCFEYGMKSERYNDNRNKKKKEVPKPECSSEYTLTPEVG 180
 DB 115 RDKNCLINKVTRNFCQYCRLOKCFEYGMKSERYNDNRNKKKKEVPKPECSSEYTLTPEVG 174
 QY 181 ELIEKVRKAHOETFPALCOLGKYYTNNSSRQVSLDLDWKFSELSTKCIITVEFAKQ 240
 DB 175 DLIEKVRKAHOETFPALCOLGKYYTNNSSRQVSLDLDWKFSELSTKCIITVEFAKQ 234
 QY 241 LFGFTTLTIADQITLLKAACLDLILRICTRYTPEDQMTFFSDGLTLNRTQMNAAGPPL 300
 DB 235 LFGFTTLTIADQITLLKAACLDLILRICTRYTPEDQMTFFSDGLTLNRTQMNAAGPPL 294
 QY 301 TLVFAFANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVDMLEALQVYVRK 360
 DB 295 TLVFAFANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVDMLEALQVYVRN 354
 QY 361 RPSRPHMFKMLMTDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEGLDTLSC 420
 DB 355 RRPKPHMFKMLMTDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEGLDTLSC 413
 QY 421 QPGGGGDDGGGLPPPGSCPSLSPSSNRSSPATHSP 457

Db 414 --GGASSQTPVTPVAPGSCSPSLSPSTHSSPSSHP 448
 RESULT 6
 Q90272
 ID Q90272 PRELIMINARY; PRT; 457 AA.
 AC Q90272;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Retinoic acid receptor alpha-2.B.
 GN RARA2B OR RAR-ALPHA-2.B.
 OS Brachydanio rerio (zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OC NCBI_TaxID=7955;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE-EMBRYO;
 RC Stachel S.E.; Kushner P.;
 RA "The molecular characterization of three zebrafish retinoic acid
 RT receptor genes suggests the retinoic acid pathway functions in
 RT embryonic hindbrain development."
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RP TISSUE-EMBRYO;
 RC Stachel S.E.;
 RA "Retinoic acid reprograms RAR expression and shows anteroposterior
 RT specification occurs in the pre-gastrula zebrafish."
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL; L03399; AAA50050.1; -.
 DR HSSP; P10826; IHRA.
 DR ZFIN; ZDB-GENE-980526-72; rara2b.
 DR InterPro; IPR000536; Hormone_rec_lig.
 DR InterPro; IPR001723; Stdhrmn_receptor.
 DR InterPro; IPR001628; Znf_C4steroid.
 DR Pfam; PF00104; hormone_rec; 1.
 DR Pfam; PF00105; zf-C4; 1.
 DR PRINTS; PR00398; STDRHORMONER.
 DR PROSITE; PS00047; STROIDFINGER.
 DR ProDom; PD000035; Znf_C4steroid; 1.
 DR SMART; SM00430; HOL1; 1.
 DR SMART; SM00399; ZNF_C4; 1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 DR DNA-binding; Nuclear protein; Receptor; Transcription regulation;
 KW Zinc-finger.
 SQ SEQUENCE 457 AA; 50481 MW; DF51E08BD6F0A625 CRC64;

 Query Match 82.9%; Score 2021; DB 13; Length 457;
 Best Local Similarity 82.4%; Pred. No. 7.4e-167;
 Matches 378; Conservative 37; Mismatches 40; Indels 4; Gaps 2;
 QY 1 MYESVEVGG--PTNPFLVDFYQNRACLLPEKGLPAPGYPSTPLRPLWNGSNHSIET 58
 DB 1 MYESVDVVGLTPSPNPFLSDYHONRGCLIPDKGLVSGA--ARGFRNPHSGSNHSVET 58
 QY 59 QSSSEIIVSPSPPPPLPRIYKPCFVCQDKSSGYHYGVSACEGCKGFFRRIQKNMVT 118
 DB 59 QTSSEIIVSPSPPPPPPRVYKPCFVCQDKSSGYHYGVSACEGCKGFFRRIQKNMVT 118
 QY 119 CHDKNCLINKVTRNFCQYCRLOKCFEYGMKSERYNDNRNKKKKEVPKPECSSEYTLTPE 178
 DB 119 CHREKSCIIINKVTRNFCQYCRLOKCFEYGMKSERYNDNRNKKKKEVPKPECSSEYTLTPE 178
 QY 179 VGLIEKVRKAHOETFPALCOLGKYYTNNSSRQVSLDLDWKFSELSTKCIITVEFA 238
 DB 179 TEKMEIQRKAHOETFPALCOLGKYYTNNSSRQVSLDLDWKFSELSTKCIITVEFA 238

```
QY 239 KOLPGFTLTATDQITLLKAAACLDLILIRICTRYPTPEQDTMTFSDGLTLNRTQMHNAFG 298
Db 239 KOLPGFTLTATDQITLLKAAACLDLILIRICTRYPTPEQDTMTFSDGLTLNRTQMHNAFG 298
QY 299 PLTDLVAFANQLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMLOEPLLEALKYVY 358
Db 299 PLTDLVAFANQLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMLOEPLLEALKYVY 358
QY 359 RKRPSRPHMPPKMLKMTDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDTL 418
Db 359 RNRPHMPPKMLKMTDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDTL 418
QY 419 SGQPGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
Db 419 GSKAGGGGGGGKAGPGSCSPSLSPSSAHSP 457

RESULT 7
Q91391 ID Q91391 PRELIMINARY; PRT; 444 AA.
AC Q91391;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Retinoic acid receptor alpha (Fragment).
GN RARA OR ZRAR-ALPHA.
OS Brachydanio rerio (zebrafish) (zebra daniel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95001557; PubMed=7918098;
RA Joore J., van der Lans G.B., Lanser P.H., Vervaaert J.M., Zivkovic D.,
RA Speksnijder J.E., Krulger W.;
RT "Effects of retinoic acid on the expression of retinoic acid receptors
RT during zebrafish embryogenesis.";
RL Mech. Dev. 46:137-150(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; S74155; AAB32276.1;
DR HSP; P10826; 1HRA.
DR ZFIN; ZDB-GENE-990415-239; rara.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrm_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 444 AA; 49924 MW; 7B0DED7B660D5E7F CRC64;
Query Match 82.4%; Score 2008.5; DB 13; Length 444;
Best Local Similarity 82.9%; Pred. No. 8.6e-166;
Matches 379; Conservatve 35; Mismatches 30; Indels 13; Gaps 3;
QY 1 MYESVEVGGTPNPFVLVDFYNACRLPEKGLPAPGPYSTPLRTPWNGSNHSIETQS 60
Db 1 MYESVDV-----NPFLLMDYNSRGLIFDK---MHPFSSSIRHWHGSGNHSIETQS 52
QY 61 SSSEIIVSPSPPLPRIYKPCFVCDKSSGYHYGSACGCKGFFRRSTQKNMYTCH 120
Db 53 TSSEIIVSPSPPLPRIYKPCFVCDKSSGYHYGSACGCKGFFRRSTQKNMYTCH 112
QY 121 RDKNCLINKYTRNQCVRQLQKCFEVMGMSKESVRNDRNKKKKEVPKPECSSESYTLTPEVG 180
```

```
Db 113 REKNCLINKYTRNQCVRQLQKCFEVMGMSKESVRNDRNKKKKEVPKPECSSESYTLTPEVG 172
QY 181 ELIEKVRKAHQETFFPALCOLGKYTTNNSEQRVSLDIDLDWDFSELSTKCIITKVEFAKQ 240
Db 173 OMIDRVKRAHQETFFSLCQLGKYTTSSNSERRVALDVLWDKFSSELSTKCIITKVEFAKQ 232
QY 241 LPEFTTLTADQITLLKAAACLDLILIRICTRYPTPEQDTMTFSDGLTLNRTQMHNAFG 300
Db 233 LPEFTTLTADQITLLKAAACLDLILIRICTRYPTPEQDTMTFSDGLTLNRTQMHNAFG 292
QY 301 TDLVAFANQLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMLOEPLLEALKYVYVRK 360
Db 293 TDLVAFANQLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMLOEPLLEALKYVYVRN 352
QY 361 RRPSPHMPKMLKMTDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDTLSG 420
Db 353 RRPSPHMPKMLKMTDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDTLSG 412
QY 421 QPGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
Db 413 AQGSRAS-----ATTFGSCSPSLSPNSAQSPPPTQSP 444

RESULT 8
Q90271 ID Q90271 PRELIMINARY; PRT; 444 AA.
AC Q90271;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Retinoic acid receptor alpha-2.A.
GN RARA2A OR ZRAR-ALPHA-2.A.
OS Brachydanio rerio (zebrafish) (zebra daniel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=EMBRYO;
RA Stachel S.E.; Kushner P.;
RT "The molecular characterization of three zebrafish retinoic acid
RT receptor genes suggests the retinoic acid pathway functions in
RT embryonic hindbrain development.";
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=EMBRYO;
RA Stachel S.E.;
RT "Retinoic acid reprograms RAR expression and shows anteroposterior
RT specification occurs in the pre-gastrula zebrafish.";
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; L03398; AAA50049.1;
DR HSP; P10826; 1HRA.
DR ZFIN; ZDB-GENE-980526-284; rara2a.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrm_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 444 AA; 49895 MW; 68380BA8460D447B CRC64;
```

Query Match 82.0%; Score 1998.5; DB 13; Length 444;
Best Local Similarity 82.5%; Pred. No. 6.4e-165;
Matches 377; Conservative 36; Mismatches 31; Indels 13; Gaps 3;

QY 1 MYESVEVGGPTNPFLVVDYFNQNRACLLPEKGLPAPGPYSTPLRTPLWNGSNHSHIETQS 60
DB 1 MYESVDV-----NPLMDYNYNGRGLIPDK---MPPHFFSSIRHQHWSGNSHSHIETQS 52
QY 61 SSSEIEIVPSPPPPPRIYKPCFCVQDKSSGYHYGVSAACGCKGFFRRSIOKNNVYTCH 120
DB 53 TSSEIEIVPSPPPPPRIYKPCFCVQDKSSGYHYGVSAACGCKGFFRRSIOKNNVYTCH 112
QY 121 RDKNCIINKVTRNCPQCYCRLOKCFEYVMSKESVNRDRNKKKKEVPKPCSESYTLTPEVG 180
DB 113 REKNCIINKVTRNCPQCYCRLOKCFEYVMSKESVNRDRNKKKKEVPKPCSESYTLTPEVG 172
QY 181 ELIEKVRKAHQETFPALCOLGKYTTNNSSEORVSLDLDWKFSELSKCIKTIKVEFAKQ 240
DB 173 QMDRVRAHQETFPALCOLGKYTTNNSSEORVSLDLDWKFSELSKCIKTIKVEFAKQ 232
QY 241 LPGFTTLTIADQITLLKAACLDIILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGL 300
DB 233 LPGFTTLTIADQITLLKAACLDIILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGL 292
QY 301 TDLVFAFANQLLPLEMDAETGLLSAICLICGDRQDLEQPDVDMQLEALKVIYVRK 360
DB 293 TDLVFAFANQLLPLEMDAETGLLSAICLICGDRQDLEQPDVDMQLEALKVIYVRK 352
QY 361 RRPSPHMFPMKIMKIIDLRSISAKGAERVITLKMIEIPGSMPLIQEMLNSEGLDTLSG 420
DB 353 RRPSPHMFPMKIMKIIDLRSISAKGAERVITLKMIEIPGSMPLIQEMLNSEGLDTLSG 412
QY 421 QPGGGGGRGGGLPPPGGSPSPSSNRSSPATHSP 457
DB 413 AQGRAS-----ATPGSCSPSPNSAQSPPTQSP 444

RESULT 9
Q8VHB7
ID Q8VHB7 PRELIMINARY; PRT; 448 AA.
AC Q8VHB7
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Retinoic acid receptor beta-2.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Sharma R.P., McGraw R.A., Dugyala R.R.;
RT "Cloning and characterization of hamster fetal retinoic acid receptor
isoforms.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046944; AAL02361.1;
DR InterPro: IPR000536; Hormone_rec.11g.
DR InterPro: IPR001723; Stdrhma_receptor.
DR InterPro: IPR001628; Znf_C4steroid.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00398; STDRHMONER.
DR PRINTS: PR00047; STROIDFINGER.
DR ProDom: PD000035; Znf_C4steroid; 1.
DR SMART: SM00430; HOLI; 1.
DR SMART: SM00399; Znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW Receptor.
SQ SEQUENCE 448 AA; 50353 MW; FC4BBEF9F7A69684 CRC64;

Query Match 72.6%; Score 1770; DB 11; Length 448;
Best Local Similarity 74.6%; Pred. No. 4.2e-145;

Matches 343; Conservative 41; Mismatches 58; Indels 18; Gaps 5;

QY 1 MYESVEVGGPTNPFLVVDYFNQNRACLLPEKGLPAPGPYSTPLRTPLWNGSN--HSIE 57
DB 1 MFDGMDVLSVSPGQ--ILDFTASPSQMLQEKALKACLSGFTQAE---WOHRHTAQISIE 55
QY 58 TOSSEIEIVPSPPPPPRIYKPCFCVQDKSSGYHYGVSAACGCKGFFRRSIOKNNVY 117
DB 56 TOSTSEIEIVPSPPPPPRIYKPCFCVQDKSSGYHYGVSAACGCKGFFRRSIOKNNVY 115
QY 118 TCHRNKCIINKVTRNCPQCYCRLOKCFEYVMSKESVNRDRNKKKKEVPKPCSESYTLTP 177
DB 116 TCHRNKNCINKVTRNCPQCYCRLOKCFEYVMSKESVNRDRNKKKKEVPKPCSESYEMTA 175
QY 178 EVGELIEKVRKAHQETFPALCOLGKYTTNNSSEORVSLDLDWKFSELSKCIKTIKVEF 237
DB 176 EDDTEKIRKAHQETFPALCOLGKYTTNNSADHRVRLDGLWDFSELSKCIKTIKVEF 235
QY 238 AKQLPGFTTLTIADQITLLKAACLDIILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAG 297
DB 236 AKRLPGFTSLTIADQITLLKAACLDIILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAG 295
QY 298 GPLTDLVFAFANQLLPLEMDAETGLLSAICLICGDRQDLEQPDVDMQLEALKVIY 357
DB 296 GPLTDLVFAFANQLLPLEMDAETGLLSAICLICGDRQDLEQPTKYDKLQEPLEALKIY 355
QY 358 VYKRSPHMFPMKIMKIIDLRSISAKGAERVITLKMIEIPGSMPLIQEMLNSEGLDT 417
DB 356 IKRRSPHMFPMKIMKIIDLRSISAKGAERVITLKMIEIPGSMPLIQEMLNSEGLHEP 415
QY 418 LSGQGGGGRGGGLPPPGGSPSPSSNRSSPATHSP 457
DB 416 LTPSSSGNTAEH-----SPSVSPSSVNSGVQSPP 445

RESULT 10
Q910C7
ID Q910C7 PRELIMINARY; PRT; 448 AA.
AC Q910C7
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Retinoic acid receptor beta-2.
GN RAR-BETA-2.
OS Coturnix coturnix (common quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=9091;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11133679;
RA Fu Z., Kubo T., Sugahara K., Noguchi T., Kato H.;
RT "Cloning of Complementary Deoxyribonucleic Acids Encoding Quail
Coturnix coturnix japonica) Retinoic Acid Receptor beta Isoforms and
Changes in their Gene Expression During Gonadotropic Growth.";
RL Biol. Reprod. 64:231-241(2001).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: AF110730; AAD23398.1;
DR InterPro: IPR000536; Hormone_rec.11g.
DR InterPro: IPR001628; Znf_C4steroid.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; zf-C4; 1.
DR ProDom: PD000035; Znf_C4steroid; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 448 AA; 50471 MW; BC8598A1D350701B CRC64;

Query Match 72.2%; Score 1760; DB 13; Length 448;
Best Local Similarity 74.6%; Pred. No. 3.1e-144;
Matches 343; Conservative 41; Mismatches 58; Indels 18; Gaps 5;

```
QY 1 MYSEVVEGGTPNFVVDYFNQ-N-RACILPEKGLPAGPYSTPLRPLM--NGSNHSIE 57
DB 1 MFCMDVLAVSAQML--DYTASPSMCLQEKALKA---CFSGLAOTWHRHSQAQSV 55
QY 58 TQSSSEIEVPSPPLPRIVKPCFVCODKSSGYHYGSACGCKGFFRRSIOKNMY 117
DB 56 TQSTSEELVPSPPLPRIVKPCFVCODKSSGYHYGSACGCKGFFRRSIOKNMY 115
QY 118 TCHRDKNCIINKYTRNPOCYCRLOKCFEVMGSKSVENDRNKKKKEVPKPECSESYTLTP 177
DB 116 TCHRDKNVINKYTRNPOCYCRLOKCFEVMGSKSVENDRNKKKKEVPKPECSESYTLTP 175
QY 178 EVELTEKVKRAHQETFPALCOQKGYTTNNSSEQRVSLDIDLMDKFSLSKCIKTVEF 237
DB 176 ELDLDEKIRKAHQETFPALCOQKGYTTNNSADHRVLDLGLWDFSELATKCIKTVEF 235
QY 238 AKOLPGFTTLTIADQITLLKAACLDILILRICRYTPEODTMTFSDGLTLNRTOMHNAGF 297
DB 236 AKRLPGFTSLTIADQITLLKAACLDILILRICRYTPEODTMTFSDGLTLNRTOMHNAGF 295
QY 298 GPLTDLVFAFANOLLPLEMDDAETGLLSAICLCGDRQDLEQDPRVDMLOEPLEALKVY 357
DB 296 GPLTDLVFAFANOLLPLEMDTETGLLSAICLCGDRQDLEQDPRVDMLOEPLEALKVY 355
QY 358 VKRRSRPHMFKMLKMTDLRSISAKGAERVITLMEIPGSMPLIQEMLENSEGLDT 417
DB 356 IRRRNKPKHMFKILMKITDLRSISAKGAERVITLMEIPGSMPLIQEMLENSEGLDT 415
QY 418 LSGOPGGGRDGGGLPPPGSCSPSLSPSNRSPATHSP 457
DB 416 LPTSTNGTAEH-----SPSISPSVDNSSVSQSP 445
```

```
RESULT 11
Q91YX2 PRELIMINARY; PRT; 458 AA.
AC Q91YX2;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Retinoic acid receptor, gamma.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; BC013709; AAH13709.1; -.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 458 AA; 50917 MW; B410D52805BE7CDD CRC64;
```

```
Query Match 67.7%; Score 1651; DB 11; Length 458;
Best Local Similarity 72.2%; Pred. No. 9.1e-135;
Matches 317; Conservative 40; Mismatches 50; Indels 32; Gaps 4;

QY 30 PEKGLPAGP-PYSTP-----LRTPLWNG-----SNHSIETQSSSEIEVP 68
DB 16 PGSGYFAGGFFAFPGALRGSPPEMLSPFRGLGQDPLPKEMASLSVETQSTSEEMVP 75
QY 69 SPPSPPLPRIVKPCFVCODKSSGYHYGSACGCKGFFRRSIOKNMYTCHRDKNCIIN 128
DB 69 SPPSPPLPRIVKPCFVCODKSSGYHYGSACGCKGFFRRSIOKNMYTCHRDKNCIIN 128
QY 129 KVTNRNPOCYCRLOKCFEVMGSKSVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 188
DB 129 KVTNRNPOCYCRLOKCFEVMGSKSVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 188
```

```
DB 76 SSPPPPPRVYKPCFVCNDKSSGYHYGVSCGCKGFFRRSIOKNMYTCHRDKNCIIN 135
QY 129 KVTNRNPOCYCRLOKCFEVMGSKSVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 188
DB 136 KVTNRNPOCYCRLOKCFEVMGSKSVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 195
QY 189 AHQETFPALCOQKGYTTNNSSEQRVSLDIDLMDKFSLSKCIKTVEFAKOLPGFTTLT 248
DB 196 AHQETFPALCOQKGYTTNNSADHRVLDLGLWDFSELATKCIKTVEFAKOLPGFTGLS 255
QY 249 IADQITLLKAACLDILILRICRYTPEODTMTFSDGLTLNRTOMHNAGFPLTDLVFAFA 308
DB 256 IADQITLLKAACLDILILRICRYTPEODTMTFSDGLTLNRTOMHNAGFPLTDLVFAFA 315
QY 309 NOLLPLEMDDAETGLLSAICLCGDRQDLEQDPRVDMLOEPLEALKVYVRRRSPRPHM 368
DB 316 GOLLPLEMDTETGLLSAICLCGDRMDLEPEKVDKLOEPLEALRYARRRRSQPYM 375
QY 369 FPKMLKITDLRSISAKGAERVITLMEIPGSMPLIQEMLENSEGLDTLSQPG----- 423
DB 376 FPKMLKITDLRSISAKGAERVITLMEIPGSMPLIQEMLENSEGLDTLSQPG----- 435
QY 424 -----GGRDGGGLPPPP 436
DB 436 SSEDEAPGGGKGKQSGPOP 454
```

```
RESULT 12
Q91VX5 PRELIMINARY; PRT; 458 AA.
AC Q91VX5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to retinoic acid receptor, gamma.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BREAST TUMOR;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; BC012923; AAH12923.1; -.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 458 AA; 50891 MW; 1D13D2F1482D7194 CRC64;
```

```
Query Match 67.7%; Score 1651; DB 11; Length 458;
Best Local Similarity 72.2%; Pred. No. 9.1e-135;
Matches 317; Conservative 40; Mismatches 50; Indels 32; Gaps 4;

QY 30 PEKGLPAGP-PYSTP-----LRTPLWNG-----SNHSIETQSSSEIEVP 68
DB 16 PGSGYFAGGFFAFPGALRGSPPEMLSPFRGLGQDPLPKEMASLSVETQSTSEEMVP 75
QY 69 SPPSPPLPRIVKPCFVCODKSSGYHYGSACGCKGFFRRSIOKNMYTCHRDKNCIIN 128
DB 76 SPPSPPLPRIVKPCFVCNDKSSGYHYGVSCGCKGFFRRSIOKNMYTCHRDKNCIIN 135
QY 129 KVTNRNPOCYCRLOKCFEVMGSKSVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 188
DB 136 KVTNRNPOCYCRLOKCFEVMGSKSVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 195
```

```
QY 189 AHOETFPALCOLGKYTNSSQVSLDLDLWDMKFSLSLTKCIKIVTFAKQLPGFTTLT 248
DB 196 AHOETFPSCQCGKYTNSSADKHYRVLQDLGLWDMKFSLSLTKCIKIVTFAKRLPGFTGLS 255
QY 249 IADQITLLKAACLDLILRICRYTPEQDTMTFSGLTLNRTOMHNGAGPLTDLVFAFA 308
DB 256 IADQITLLKAACLDLILRICRYTPEQDTMTFSGLTLNRTOMHNGAGPLTDLVFAFA 315
QY 309 NOLLPLEMDAETGLLSAICLCGRDQDLEQDPRVDMLEQLEALKVYVRRRPSRPHM 368
DB 316 GOLLPLEMDTETGLLSAICLCGRDMDLEPEKVDKLEQLEALRLYARRRRSOPYM 375
QY 369 PPKMLKTLDRSISAKGAERVITLKMETPGSMPLLOEMNSEGLDLSQPG----- 423
DB 376 PPKMLKTLDRSISAKGAERVITLKMETPGSMPLLOEMNSEGLDLSQPG----- 423
QY 424 -----GGGRDGGGLPPPP 436
DB 436 SDEAPGCGQKRGQSQPP 454

RESULT 13
QY 08VHB6 PRELIMINARY; PRT; 443 AA.
AC 08VHB6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Retinoic acid receptor gamma-2.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Sharma R.P., McGraw R.A., Dugyala R.R.;
RT "Cloning and characterization of hamster fetal retinoic acid receptor
isoforms."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046945; AL023621;
DR InterPro: IPR00536; Hormone_rec_llg.
DR InterPro: IPR001723; Stdhrmn_receptor.
DR InterPro: IPR001628; Znf_C4steroid.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00398; STRDHORMONER.
DR PRINTS: PR00047; STROIDFINGER.
DR ProDom: PD000035; Znf_C4steroid; 1.
DR SMART: SM00430; HOLI; 1.
DR SMART: SM00399; Znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW Receptor.
SQ SEQUENCE 443 AA; 49431 MW; 36D86F26CBC4AB49 CRC64;

Query Match 67.6%; Score 1646.5; DB 11; Length 443;
Best Local Similarity 79.3%; Pred. No. 2.1e-134;
Matches 314; Conservative 32; Mismatches 37; Indels 13; Gaps 3;

QY 52 SNHSIETQSSSEIIVPSPPLPRIYKPCFVCODKSSGYHYGVSAECGCKGFFRRSI 111
DB 48 SLOSVETQSTSSSEIIVPSPPLPRIYKPCFVCNCKSSGYHYGVSSCEGCKGFFRSI 107
QY 112 QKNWYTCRDKNCIINKVTRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCS- 170
DB 108 QKNWYTCRDKNCIINKVTRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEV-KEECSP 166
QY 171 ESYTLTPEVGLIEKVRKAHQETFPALCOLGKYTNSSQVSLDLDLWDMKFSLSLTK 230
DB 167 DSYELSPQLEELTKYKAHQETFPALCOLGKYTNSSADKHYRVLQDLGLWDMKFSLSLTK 226
QY 231 IIKTVFAKQLPGFTTLTADQITLLKAACLDLILRICRYTPEQDTMTFSGLTLNRT 290
DB 231 IIKTVFAKQLPGFTTLTADQITLLKAACLDLILRICRYTPEQDTMTFSGLTLNRT 290
```

```
DB 227 IIKVIFAKRLPGFTGLSIADQITLLKAACLDLILMLRICRYTPEQDTMTFSGLTLNRT 286
QY 291 QHNAGFGPLTDLVFAFANOLLPLEMDAETGLLSAICLCGRDQDLEQDPRVDMLEQPL 350
DB 287 QHNAGFGPLTDLVFAFAGOLLPLEMDTETGLLSAICLCGRDMDLEPEKVDKLEQPL 346
QY 351 LEALKVYVRRRPSRPHMFKMLKTLDRSISAKGAERVITLKMETPGSMPLLOEMLE 410
DB 347 LEALRLYARRRRSOPYMFKMLKTLDRSISAKGAERVITLKMETPGSMPLLOEMLE 406
QY 411 NSEGLDLSQPG-----GGGRDGGGLPPP 435
DB 407 NPEMFEDDSKSGPHPKASSEDETPGQKGVGLKPP 442
QY 411 NSEGLDLSQPG-----GGGRDGGGLPPP 435
DB 407 NPEMFEDDSKSGPHPKASSEDETPGQKGVGLKPP 442

RESULT 14
QY 0918T3 PRELIMINARY; PRT; 441 AA.
AC 0918T3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Retinoic acid receptor gamma 2.
GN NR1B3B.
OS Ambystoma mexicanum (Axolotl).
OC Plasmid araRgamma2.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidae; Ambystomatidae;
OC Ambystoma.
OX NCBI_TaxID=8296;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE=20461762; PubMed=11004482;
RA Wirtanen L., Seguin C.;
RT "Cloning of cDNAs encoding retinoic acid receptors RARGamma1,
RARGamma2, and a new splicing variant, RARGamma3, from Ambystoma
mexicanum and characterization of their expression during early
development."
RL Blochm. Biophys. Acta 1492:81-93(2000).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: AF206315; AAF80975.1;
DR HSSP: P22932; 3LBD.
DR InterPro: IPR000536; Hormone_rec_llg.
DR InterPro: IPR001723; Stdhrmn_receptor.
DR InterPro: IPR001628; Znf_C4steroid.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00398; STRDHORMONER.
DR PRINTS: PR00047; STROIDFINGER.
DR ProDom: PD000035; Znf_C4steroid; 1.
DR SMART: SM00430; HOLI; 1.
DR SMART: SM00399; Znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
KW DNA-binding; Nuclear protein; Plasmid; Receptor;
KW Transcription regulation; Zinc-finger.
SQ SEQUENCE 441 AA; 49865 MW; 906E25E145CEESE CRC64;
```

```
Query Match 67.3%; Score 1641; DB 13; Length 441;
Best Local Similarity 74.8%; Pred. No. 6.4e-134;
Matches 318; Conservative 37; Mismatches 56; Indels 14; Gaps 6;

QY 1 MYESVEGGTTPNPFVVDYFNQNRACILLPEKGL-PAPG---PYSTPLRTPLWNGSNHSI 56
DB 1 MYDCMEAFMLAPHP--LYDVTNPG-ACMLRKLSPCFGLDPFGWFPASL-----QSV 52
QY 57 ETQSSSEIIVPSPPLPRIYKPCFVCODKSSGYHYGVSAECGCKGFFRRSIQKNV 116
DB 53 ETQSSSEIIVPSPPLPRIYKPCFVCNCKSSGYHYGVSSCEGCKGFFRRSIQKNV 112
QY 117 YTCRDKNCIINKVTRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCSSEYTLT 176
DB 117 YTCRDKNCIINKVTRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCSSEYTLT 176
```

Search completed: March 29, 2003, 07:57:21
Job time : 65 secs

Query Match	66.8%;	Score 1628.5;	DB 13;	Length 446;
Best Local Similarity	74.0%;	Pred. No. 7.9e-133;		
Conservative	37.	Mismatches	56;	Indels 19;
W-L-L-L	318.	Gaps	7;	

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: March 29, 2003, 05:12:23 ; Search time 23 Seconds
(without alignments)

824.117 Million cell updates/sec

Title: US-09-691-220-2

Perfect score: 2437

Sequence: 1 MYSEVEGPTPNPLVDF.....SCSPSLSPSSNRSSPATHSP 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2127	87.3	462	1 RRA_HUMAN	P10276 homo sapien
2	2089	85.7	462	1 RRA_MOUSE	P11416 mus musculus
3	2020.5	82.9	460	1 RRA_CHICK	Q90966 gallus gall
4	1972	80.9	458	1 RRA_NOTVI	P18514 notophthalm
5	1895	77.8	458	1 RRA_XENLA	P51126 xenopus lae
6	1868.5	76.7	447	1 RRA_FUGRU	Q9W323 fugu rubrip
7	1761	72.3	455	1 RRB_HUMAN	P10826 homo sapien
8	1753.5	72.0	482	1 RRB_MOUSE	P22605 mus musculus
9	1736	71.2	455	1 RRB_COTJA	Q9W6b3 coturnix co
10	1733	71.1	455	1 RRB_CHICK	P22448 gallus gall
11	1663.5	68.3	499	1 RRG_BRARE	Q91392 brachydanio
12	1662	68.2	454	1 RRG1_HUMAN	P13631 homo sapien
13	1655.5	67.9	443	1 RRG2_HUMAN	P22932 homo sapien
14	1642	67.4	458	1 RRG1_MOUSE	P18911 mus musculus
15	1633.5	67.0	447	1 RRG2_MOUSE	P20787 mus musculus
16	1627.5	66.8	442	1 RRG2_XENLA	P28699 xenopus lae
17	1621	66.5	505	1 RRG_NOTVI	P18516 notophthalm
18	1620.5	66.5	476	1 RRG1_XENLA	P51127 xenopus lae
19	724	29.7	158	1 RRB_NOTVI	P18515 notophthalm
20	676.5	27.8	427	1 THA1_BRARE	Q98867 brachydanio
21	659.5	27.1	418	1 THAA_XENLA	P15204 xenopus lae
22	658	27.0	418	1 THAB_XENLA	P18115 xenopus lae
23	657.5	27.0	416	1 THA_SALSA	Q9W785 salmo salar
24	653	26.8	410	1 THAI_SHEEP	Q28570 ovis aries
25	647	26.5	416	1 THA_HIPHI	Q9W6n4 hippoglossu
26	647	26.5	418	1 THA_RANCA	Q02777 rana catesb
27	646.5	26.5	391	1 THAB_PAROL	Q91242 paralichthy
28	646	26.5	416	1 THAA_PAROL	Q91241 paralichthy
29	645.5	26.5	408	1 THA_CHICK	P04625 gallus gall
30	639	26.2	373	1 THB_RANCA	Q02965 rana catesb
31	634.5	26.0	401	1 THA_CAIMO	Q90382 cairina mos
32	630.5	25.9	395	1 THB_PAROL	Q91279 paralichthy
33	630	25.9	373	1 THBA_XENLA	P18117 xenopus lae

34 629.5 25.8 402 1 THA_APTPA
35 629 25.8 386 1 THB_BRARE
36 625.5 25.7 475 1 THB2_MOUSE
37 625 25.6 414 1 THBB_XENLA
38 622.5 25.5 461 1 THB1_HUMAN
39 619 25.4 514 1 THB2_RAT
40 618 25.4 369 1 THB2_CHICK
41 617 25.3 476 1 THB2_HUMAN
42 615 25.2 461 1 THB1_MOUSE
43 615 25.2 461 1 THB1_RAT
44 607.5 24.9 402 1 THA_PYGAD
45 606.5 24.9 385 1 ERBA_AVIER

ALIGNMENTS

RESULT 1

RRA_HUMAN STANDARD; PRT: 462 AA.
AC P10276; Q13440; Q13441; P78456; Q9NQS0;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retinoic acid receptor alpha (RAR-alpha).
GN RARA OR NR1B1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA-1).
RX MEDLINE=88055922; PubMed=2825036;
RT Giguere V., Ong E.S., Segui P., Evans R.M.;
RA "Identification of a receptor for the morphogen retinoic acid.";
RL Nature 330:624-629(1987).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA-1).
RX MEDLINE=941134721; PubMed=8302850;
RA Chen Z., Guidet F., Rousselot P., Agadir A., Chen S.-J., Wang Z.-Y.,
Degos L., Zalent A., Waxman S., Chomienne C.;
RT "PLZF-RAR alpha fusion proteins generated from the variant
t(11;17)(q23;q21) translocation in acute promyelocytic leukemia
inhibit ligand-dependent transactivation of wild-type retinoic acid
receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:1178-1182(1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA-1).
RX MEDLINE=99269925; PubMed=10337631;
RA Hjalt T.A.H., Murray J.C.;
RT "Genomic structure of the human retinoic acid receptor-alpha gene.";
RL Mamm. Genome 10:528-529(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA-1).
RX TISSUE=Brain;
RA Strausberg R.;
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-80 FROM N.A. (ISOFORM ALPHA-1).
RX MEDLINE=91088249; PubMed=2175878;
RA Brand N.J., Petkovich M., Chambon P.;
RT "Characterization of a functional promoter for the human retinoic acid
receptor-alpha (hRAR-alpha).";
RL Nucleic Acids Res. 18:6799-6806(1990).
RN [6]
RP SEQUENCE OF 31-462 FROM N.A.
RX MEDLINE=88055872; PubMed=2825025;
RA Petkovich M., Brand N.J., Krust A., Chambon P.;
RT "A human retinoic acid receptor which belongs to the family of nuclear
receptors.";
RL Nature 330:444-450(1987).
RN [7]
RP REVISIONS.

RA Zelent A., Krust A., Petkovitch M., Kastner P., Chambon P.;
 RT "Cloning of murine alpha and beta retinoic acid receptors and a novel
 RL receptor gamma predominantly expressed in skin.";
 RN Nature 339:714-717(1989).
 RP [2]
 RX SEQUENCE FROM N.A. (ISOFORM ALPHA-1).
 RA MEDLINE-93195958; PubMed-8383767;
 RA Heilmann R., Rentrop M., Lang E., Maelicke A.;
 RT "Cloning of several genes coding for retinoic acid nuclear receptors
 RL in the mouse embryonal carcinoma cell line PCC7-MZ1.";
 RN J. Recept. Res. 13:693-709(1993).
 RP [3]
 RX SEQUENCE FROM N.A. (ISOFORMS ALPHA-1 AND ALPHA-2).
 RA MEDLINE-91114713; PubMed-1846598;
 RA Leroy P., Krust A., Zelent A., Mendelsohn C., Garnier J.-M.,
 RA Kastner P., Dierich A., Chambon P.;
 RT "Multiple isoforms of the mouse retinoic acid receptor alpha are
 RL generated by alternative splicing and differential induction by
 RL retinoic acid.";
 RN EMBO J. 10:59-69(1991).
 RP [4]
 RX SEQUENCE FROM N.A. (VARIANT IN EMBRYONAL CARCINOMA CELL LINE RAC65).
 RA MEDLINE-92324516; PubMed-1320576;
 RA Krut F.A.E., van der Veer L., Mader S., van den Brink C.E.,
 RA Feijen A., Jonk L.J., Kruijer W., van der Saag P.T.;
 RT "Retinoic acid resistance of the variant embryonal carcinoma cell
 RL line RAC65 is caused by expression of a truncated RAR alpha.";
 RN Differentiation 49:27-37(1992).
 RP [5]
 RX SEQUENCE FROM N.A. (VARIANT IN EMBRYONAL CARCINOMA CELL LINE RAC65).
 RA MEDLINE-91061752; PubMed-2174108;
 RA Pratt M.A.C., Kralova J., McBurney M.W.;
 RT "A dominant negative mutation of the alpha retinoic acid receptor
 RL gene in a retinoic acid-nonresponsive embryonal carcinoma cell.";
 RN Mol. Cell. Biol. 10:6445-6453(1990).
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE
 CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
 CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
 CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA-1 (SHOWN HERE) AND ALPHA-
 CC 2; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
 CC NR1 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X56572; CAA39919.1; -;
 DR EMBL; X56565; CAA39917.1; -;
 DR EMBL; S56656; AAB25783.1; -;
 DR EMBL; X57528; CAA40749.1; -;
 DR EMBL; M60909; AAA40031.1; -;
 DR PIR; S05050; S05050.
 DR PIR; S13599; S13599.
 DR HSP; P10826; 1HRA.
 DR TRANSFAC; T00717; -;
 DR TRANSFAC; T01327; -;
 DR TRANSFAC; T01335; -;
 DR TRANSFAC; T01336; -;
 DR MGD; MGI:97856; Rara.
 DR InterPro; IPR000536; Hormone_rec_lig.
 DR InterPro; IPR001723; Steroid_receptor.
 DR InterPro; IPR001628; Znf_C4steroid.
 DR Pfam; PF00104; hormone_rec; 1.
 DR Pfam; PF00105; zf-C4; 1.

DR PRINTS; PR00398; STRDHORMONER.
 DR PRINTS; PR00047; STROIDFINGER.
 DR PRODOM; PD000035; Znf_C4steroid; 1.
 DR SMART; SM00430; HOL1; 1.
 DR SMART; SM00399; Znf_C4; 1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 RL Zinc-finger; Multigene family; Alternative splicing.
 FT DOMAIN 1 87 MODULATING.
 FT DNA_BIND 88 153 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 88 108 C4-TYPE.
 FT ZN_FING 124 148 C4-TYPE.
 FT DOMAIN 154 199 HINGE.
 FT DOMAIN 200 419 LIGAND-BINDING.
 FT VARSPLIC 1 60
 FT VARIANT 391 391 MISSING (IN EMBRYONAL CARCINOMA CELL LINE
 FT VARIANT 392 462 RAC65).
 FT CONFLICT 163 163 N -> K (IN REF. 5).
 FT CONFLICT 179 179 T -> S (IN REF. 5).
 FT CONFLICT 284 284 M -> L (IN REF. 5).
 SQ SEQUENCE 462 AA; 50735 MW; 726F77996338A5AD CRC64;
 Query Match 85.7%; Score 2089; DB 1; Length 462;
 Best Local Similarity 90.0%; Pred. No. 1.6e-146;
 Matches 406; Conservative 8; Mismatches 17; Indels 20; Gaps 4;
 QY 7 VGGTPNPFLVVDYNNRACLLPEKGLPAGPYSTPLRTPLWNGSNHSIETQSSSEI 66
 DB 32 LGGLSLPGALTSLQHQ-----LPVSG-YSTP-----SPATIETQSSSEI 71
 QY 67 VPSPSPPLPRIYKPCFVCDKSSGVHYGVSAEGCKGKGFERRSIQKNVYTCHRDKNCI 126
 DB 72 VPSPSPPLPRIYKPCFVCDKSSGVHYGVSAEGCKGKGFERRSIQKNVYTCHRDKNCI 131
 QY 127 INKVTNRPCQYCRLOKCFEVMGSKESVNRDRNKKKKYKPKCESYTLTPVEGELIEKV 186
 DB 132 INKVTNRPCQYCRLOKCFEVMGSKESVNRDRNKKKKYKPKCESYTLTPVEGELIEKV 191
 QY 187 RKAHQETFPALCOLGKYTTNNSSQQRVSLDIDLMDKFSLSKCIKTVEFAKQLPGFTT 246
 DB 192 RKAHQETFPALCOLGKYTTNNSSQQRVSLDIDLMDKFSLSKCIKTVEFAKQLPGFTT 251
 QY 247 LTIADQITLLKAACLDILILICITRYTPEQDTMTFSDGLTLNRTOMHNAFGPLTDLVFA 306
 DB 252 LTIADQITLLKAACLDILILICITRYTPEQDTMTFSDGLTLNRTOMHNAFGPLTDLVFA 311
 QY 307 FANOLLPLEMDDAETGLLSAICLCIGDRQDLQPRVDMQLQEPLEALKVYVKKRPSRP 366
 DB 312 FANOLLPLEMDDAETGLLSAICLCIGDRQDLQPRVDMQLQEPLEALKVYVKKRPSRP 371
 QY 367 HMFPEKMLKMTDLRSISAKGAERVITLKMETPGSNMPLIQEMLENSEGLDTLSGPGGG 426
 DB 372 HMFPEKMLKMTDLRSISAKGAERVITLKMETPGSNMPLIQEMLENSEGLDTLSGPGGG 431
 QY 427 RDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
 DB 432 RDGGGLAPPPGSCSPSLSPSSNRSSPATHSP 462
 RESULT 3
 ID RRA_CHICK STANDARD; PRT; 460 AA.
 AC Q90966; Q90967;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Retinoic acid receptor alpha (RAR-alpha).
 GN RARA OR NR1B1.

QY 116 VYCHRDKNKCIINKVTRNRCQYCLQKCFEVMGSKESVRNDRNKKKVEPKPECSYTL 175
Db 121 VYCHRDKNKCIINKVTRNRCQYCLQKCFEVMGSKESVRNDRNKKKVEPKPEAESYIL 180
QY 176 TPEVGLIEKVRKAHQETFFALCOLGKYTYNNSSQVRSLDLDWDFSELSSTCKIITV 235
Db 181 SPETQDUIEKVKAHQETFFALCOLGKYTYNSQVRSLDLDWDFSELSSTCKIITV 240
QY 236 EFAKQLPGFTTLTIADQITLLKAACTLLILRICRYTTPEDTWTFSDDLTLNRTQMHNA 295
Db 241 EFAKQLPGFTTLTIADQITLLKAACTLLILRICRYTTPEDTWTFSDDLTLNRTQMHNA 300
QY 296 GFGLDLYFAFANQLPLEMDAETGLLSAICLICGDRDLEOPDRVDMQLQEPLEALK 355
Db 301 GFGLDLYFAFANQLPLEMDAETGLLSAICLICGDRDLEOPDRVDMQLQEPLEALK 360
QY 356 VYVYKRRPSRPHFPMKMLKITDLSISAKGAERVITLKEIPEGMPPLTQEMLENSEGL 415
Db 361 IYVTRRPQKPHFPMKMLKITDLSISAKGAERVITLKEIPEGMPPLTQEMLENSEGL 419
QY 416 DTLSGQPGGGRGGGLPPPPGSCPSLSPSSNRSSPATHSP 457
Db 420 DTLSGQPGGGRGGGLPPPPGSCPSLSPSSNRSSPATHSP 458

RESULT 6

RRR_FUGRU
ID RRA_FUGRU STANDARD; PRT; 447 AA.
AC QW523; QW524;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retinoic acid receptor alpha (RAR-alpha).
GN RARA OR NR1B1.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99384012; PubMed=10452951;
RA Wentworth J.M., Schoenfeld V., Meek S., Elgar G., Brenner S.,
R Chatterjee K.K.;
RT "Isolation and characterization of the retinoic acid receptor-alpha
gene in the Japanese pufferfish, F. rubripes.";
RL Gene 236:315-323(1999).
CC -!- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE
HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA-1 (SHOWN HERE) AND ALPHA-
2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR1 SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AJ012382; CAB96754.1; -
DR EMBL; AJ012380; CAB96754.1; JOINED.
DR EMBL; AJ012381; CAB43979.1; -
DR EMBL; AJ012380; CAB43979.1; JOINED.
DR EMBL; AJ012378; CAB43870.1; -

DR EMBL; AJ012379; CAB43871.1; -
DR HSP; P10826; IHRA.
DR InterPro; IPR00536; Hormone_rec_lig.
DR InterPro; IPR001626; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 2.
DR Pfam; PF00105; zf-C4; 2.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD00035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family; Alternative splicing.
FT DOMAIN 1 79
FT DNA_BIND 80 145
FT ZN_FING 80 100
FT ZN_FING 116 140
FT DOMAIN 146 191
FT DOMAIN 192 411
FT VARSPLIC 1 52
FT FT
FT FT
FT FT
SQ SEQUENCE 447 AA; 49532 MW; E00630F720B1508D CRC64;
Query Match 76.7%; Score 1868.5; DB 1; Length 447;
Best Local Similarity 79.7%; Pred. No. 2.6e-130;
Matches 366; Conservative 25; Mismatches 43; Indels 25; Gaps 7;
QY 8 GGTPNPFL-----VDFYVQNRACLLPE--KGLPAGPYSTPLRTPLMNGSNHSIET 58
Db 5 GNPVPGPHLNGFPVPTYSYFPPHMLGSLSPALPGLISG-YSTP-----SPATLET 55
QY 59 QSSSEIEIVSPSPPLPRIYKPCFVCQDKSSGHHYGVSAECGCKGFFRSIQKNMYT 118
Db 56 QSTSEIEIVSPSPPPPPRVRKYKPCFVCQDKSSGHHYGVSAECGCKGFFRSIQKNMYT 115
QY 119 CHRDKNCIINKVTRNRCQYCLQKCFEVMGSKESVRNDRNKKKVEPKPECSYTLAPE 178
Db 116 CHREKNCIINKVTRNRCQYCLQKCFEVMGSKESVRNDRNKKKVEPKPECIENVLSPD 175
QY 179 VGLIEKVRKAHQETFFALCOLGKYTYNNSSQVRSLDLDWDFSELSSTCKIITVEFA 238
Db 176 TEQMINVRKAHQETFFALCOLGKYTYNNSSQVRSLDLDWDFSELSSTCKIITVEFA 235
QY 239 KQLPGFTTLTIADQITLLKAACTLLILRICRYTTPEDTWTFSDDLTLNRTQMHNAFG 298
Db 236 KQLPGFTTLTIADQITLLKAACTLLILRICRYTTPEDTWTFSDDLTLNRTQMHNAFG 295
QY 299 PLTDVFAFANQLPLEMDAETGLLSAICLICGDRDLEOPDRVDMQLQEPLEALKYV 358
Db 296 PLTDVFAFANQLPLEMDAETGLLSAICLICGDRDLEOPDRVDMQLQEPLEALKYV 355
QY 359 RKRPSRPHFPMKMLKITDLSISAKGAERVITLKEIPEGMPPLTQEMLENSEGLDL 418
Db 356 RRRPHFPMKMLKITDLSISAKGAERVITLKEIPEGMPPLTQEMLENSEGL-- 413
QY 419 SGQPGGGRGGGLPPPPGSCPSLSPSSNRSSPATHSP 457
Db 414 SG--ATGSRPSGA---PPGSCPSLSPSSAQSPPPTQSP 447

RESULT 7

RRR_HUMAN
ID RRB_HUMAN STANDARD; PRT; 455 AA.
AC P10826; Q00989; Q15298; Q9UN48; P12891;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retinoic acid receptor beta (RAR-beta) (RAR-epsilon) (HBV-activated
protein).
GN RARB OR NR1B2 OR HAP.
OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM BETA-2).
 RC TISSUE=Placenta;
 RX MEDLINE=88232961; PubMed=2836738;
 RA Benbrook D., Lernherdt E., Pfahl M.;
 RT "A new retinoic acid receptor identified from a hepatocellular
 RL carcinoma.";
 RL Nature 333:669-672(1988).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM BETA-2).
 RX MEDLINE=88065931; PubMed=2825037;
 RA de The H., Marchio A., Tiollais P., Dejean A.;
 RT "A novel steroid thyroid hormone receptor-related gene
 RL inappropriately expressed in human hepatocellular carcinoma.";
 RL Nature 330:667-670(1987).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM BETA-4).
 RC TISSUE=Breast tumor;
 RX MEDLINE=99342073; PubMed=10411930;
 RA Sommer K.M., Chen L.I., Treuting P.M., Smith L.T., Swisshelm K.;
 RT "Elevated retinoic acid receptor beta(4) protein in human breast tumor
 RL cells with nuclear and cytoplasmic localization.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:8651-8656(1999).
 RN [4]
 RP SEQUENCE OF 1-72 FROM N.A. (ISOFORM BETA-1).
 RX MEDLINE=94101604; PubMed=8275470;
 RA Houle B., Pelletier M., Wu J., Goddard C., Bradley W.E.;
 RT "Petal isoform of human retinoic acid receptor beta expressed in small
 RL cell lung cancer lines.";
 RL Cancer Res. 54:365-369(1994).
 RN [5]
 RP SEQUENCE OF 61-109 FROM N.A.
 RX MEDLINE=86257411; PubMed=3014347;
 RA Dejean A., Bouguieret L., Grzeschik K.-H., Tiollais P.;
 RT "Hepatitis B virus DNA integration in a sequence homologous to
 RL v-erbA and steroid receptor genes in a hepatocellular carcinoma.";
 RL Nature 322:70-72(1986).
 RN [6]
 RP SEQUENCE OF 61-109 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=91014673; PubMed=2170809;
 RA Dejean A., de The H.;
 RT "Hepatitis B virus as an insertional mutagen in a human
 RL hepatocellular carcinoma.";
 RL Mol. Biol. Med. 7:213-222(1990).
 RN [7]
 RP IDENTITY OF HAP AND RAR-BETA.
 RX MEDLINE=88189352; PubMed=2833708;
 RA Brand N., Petkovitch M., Krust A., Chambon P., de The H., Marchio A.,
 RT Tiollais P., Dejean A.;
 RL "Identification of a second human retinoic acid receptor.";
 RL Nature 332:850-853(1988).
 RN [8]
 RP STRUCTURE BY NMR OF 82-160.
 RX MEDLINE=92338160; PubMed=1321662;
 RA Kathira M., Knegetel R.M.A., Boelens R., Eib D., Schilthuis J.G.,
 RT van der Saag P.T., Kaptein R.;
 RL "Homo- and heteronuclear NMR studies of the human retinoic acid
 RT receptor beta DNA-binding domain: sequential assignments and
 RL identification of secondary structure elements.";
 RL Biochemistry 31:6474-6480(1992).
 RN [9]
 RP STRUCTURE BY NMR OF 82-160.
 RX MEDLINE=93192652; PubMed=8383553;
 RA Knegetel R.M., Kathira M., Schilthuis J.G., Bonvin A.M., Boelens R.,
 RT Eib D., van der Saag P.T., Kaptein R.;
 RL "The solution structure of the human retinoic acid receptor-beta DNA-
 RT binding domain.";
 RL J. Biomol. NMR 3:1-17(1993).
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE

CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
 CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
 CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (isoforms beta-1 and beta-2) and
 CC cytoplasmic (isoform beta-4).
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; beta-1 (shown here), beta-2,
 CC beta-3 and beta-4; are produced by alternative splicing.
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
 CC NR1 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X07282; CAA30262.1; -
 CC EMBL; Y00291; CAA68398.1; -
 CC EMBL; AF157483; RAD45688.1; -
 CC EMBL; X77664; CAA54740.1; -
 CC EMBL; X04014; CAA27637.1; ALT_SEQ.
 CC EMBL; M57445; AAA58728.1; -
 CC PIR; S02827; S02827.
 CC PIR; A29492; A29492.
 CC PIR; A25721; TVHUEA.
 CC PDB; 1HRA; 31-JAN-94.
 CC TRANSFAC; T00721; -
 CC TRANSFAC; T01326; -
 CC Gene; HGNC:9865; RARB.
 CC MIM; 180220; -
 CC InterPro; IPR000536; Hormone_rec_lig.
 CC InterPro; IPR001723; Steroid_receptor.
 CC InterPro; IPR001628; Znf_C4steroid.
 CC Pfam; PF00104; hormone_rec; 2.
 CC Pfam; PF00105; zf-C4; 4.
 CC PRINTS; PR00398; STRDHORMONER.
 CC PROSITE; PR00047; STEROIDFINGER.
 CC PROSITE; PR000035; Znf_C4steroid; 1.
 CC SMART; SM00430; HOL1; 2.
 CC SMART; SM00399; Znf_C4; 2.
 CC PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 CC Zinc-finger; Multigene family; Alternative splicing; Proto-oncogene;
 CC 3D-structure.
 CC DOMAIN 1 87
 CC DNA_BIND 88 153
 CC ZN_FING 88 108
 CC ZN_FING 124 148
 CC DOMAIN 154 199
 CC DOMAIN 200 419
 CC VARSPLIC 1 60
 CC MODULATING.
 CC NUCLEAR RECEPTOR-TYPE.
 CC C4-TYPE.
 CC C4-TYPE.
 CC HINGE.
 CC LIGAND-BINDING.
 CC MTTSGHACVPVAVNGHMTHTYPATYPLLEPPVIGLSLPL
 CC HGLHGHPPSGGCTSPAT -> MFCMDVLSVSPGQILDF
 CC YTASPSMLOKALKACFSGLTQTEWQHRHTAQS (IN
 CC ISOFORM BETA-2).
 CC MISSING (IN ISOFORM BETA-4).
 CC G -> A (IN REF. 2).
 CC L -> Q (IN REF. 1).
 CC L -> M (IN REF. 2).
 CC V -> L (IN REF. 1).
 CC
 CC VARSPLIC 1 119
 CC CONFLICT 206 206
 CC CONFLICT 317 317
 CC CONFLICT 414 414
 CC CONFLICT 454 454
 CC HELIX 107 118
 CC TURN 119 119
 CC TURN 128 129
 CC HELIX 141 151
 CC TURN 158 159
 CC SEQUENCE 455 AA; 50489 MW; 8813263AD0495D5A CRC64;
 CC
 CC Query Match 72.3%; Score 1761; DB 1; Length 455;
 CC Best Local Similarity 79.1%; Pred. NO. 2.2e-122;
 CC Matches 336; Conservative 30; Mismatches 41; Indels 18; Gaps 2;

```
QY 33 GLPAPGYPSTPLRTPLWNGSNHSIETQSSSEBIVSPSPPLPRIYKPCFVCQDKSSG 92
Db 46 GHPPPSGCSTP-----SPATIEQTSSSELVSPSPPLPPRVKPCFVCQDKSSG 97
QY 93 YHGVYACGCGKGFRRSIQKNVYTCHRDKNCLINKVTRNCPQYCRLOKCFEFGVSKES 152
Db 98 YHGVYACGCGKGFRRSIQKNVYTCHRDKNCLINKVTRNCPQYCRLOKCFEFGVSKES 157
QY 153 VRNDRNKKKKVPKPCSSSYTLTPVEGBELIEKVRKHAQETFPALCQLGKYTTNNSSEOR 212
Db 158 VRNDRNKKKKETSKQECTSEYEMTAEELDLTEKIRKHAQETFPSCQLGKYTTNNSADHR 217
QY 213 VSLDLDLWDFSELSTKCIKIIVFAKQLPGFTTLTIADQITLLKACLDLILRLICTRY 272
Db 218 VRLDGLWDFSELATKCIKIIVFAKRLPGFTTLTIADQITLLKACLDLILRLICTRY 277
QY 273 TPEQDTMTFSDGLTLNRTOMHNGFGLDLVAFANOLLPLEMDDAETGLLSAICLICG 332
Db 278 TPEQDTMTFSDGLTLNRTOMHNGFGLDLVFTFANQLLPLEMDDETGLLSAICLICG 337
QY 333 DRQDLQPDVMDLQEPLEALKVYKRRSPRPHMFKMLMKITDLRSISAKGAERVIT 392
Db 338 DRQDLQEPKVKLOEPLEALKVYKRRSPKPHMFKMLMKITDLRSISAKGAERVIT 397
QY 393 LKMEIPGSMPLIQEMLNSEGDLTSLSGQGGGGGLPPPGSCSPSLSPSSNRSSP 452
Db 398 LKMEIPGSMPLIQEMLNSEGHEPLTPSSSGNTAEH-----SPSISPSSVENS 447
QY 453 ATHSP 457
Db 448 VSQSP 452

RESULT 8
RBL_MOUSE
ID RBL_MOUSE STANDARD; PRT; 482 AA.
AC P22605; P22604; P11417;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retinoic acid receptor beta (RAR-beta).
GN RARB OR NR1B2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS BETA-1; BETA-2 AND BETA-3).
RC STRAIN=C57BL/6;
RX MEDLINE=91114714; PubMed=1846599;
RA Zelen A., Mendelsohn C., Kastner P., Krust A., Garnier J.-M.,
RA Ruffenach F., Leroy P., Chambon P.;
RT "Differentially expressed isoforms of the mouse retinoic acid
RT receptor beta generated by usage of two promoters and alternative
RT splicing.";
RL EMBO J. 10:71-81(1991).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM BETA-2).
RX MEDLINE=89295563; PubMed=2544807;
RA Zelen A., Krust A., Petkovitch M., Kastner P., Chambon P.;
RT "Cloning of murine alpha and beta retinoic acid receptors and a novel
RT receptor gamma predominantly expressed in skin.";
RL Nature 339:714-717(1989).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM BETA-2).
RX MEDLINE=93195856; PubMed=8383767;
RA Heiermann R., Rentrop M., Lang E., Maelicke A.;
RT "Cloning of several genes coding for retinoic acid nuclear receptors
RT in the mouse embryonal carcinoma cell line PCC7-M21.";
RL J. Recept. Res. 13:693-709(1993).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM BETA-4).
```

```
RX MEDLINE=92212900; PubMed=1313565;
RA Nagpal S., Zient A., Chambon P.;
RT "RAR-beta 4, a retinoic acid receptor isoform is generated from
RT RAR-beta 2 by alternative splicing and usage of a CUG initiator
RT codon.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:2718-2722(1992).
CC -1- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE
CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; BETA-1, BETA-2, BETA-3 (SHOWN
CC HERE) AND BETA-4; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NRL SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56574; CAA39921.1; -
DR EMBL; X56569; CAA39918.1; -
DR EMBL; X56573; CAA39920.1; -
DR EMBL; S56660; AAB25784.2; -
DR EMBL; S92180; -; NOT_ANNOTATED_CDS.
DR PIR; S13515; S13515
DR PIR; S13516; S13516
DR PIR; S13517; S13517
DR PIR; S05051; S05051
DR HSP; P10826; 1HRA.
DR TRANSFAC; T00717; -
DR TRANSFAC; T01337; -
DR TRANSFAC; T01338; -
DR TRANSFAC; T01339; -
MGD; MGI:97857; Rarb.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrnm_receptor.
DR Pfam; PF00104; Znf_C4steroid.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR PRODom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family; Alternative splicing.
FT DOMAIN 1 114 MODULATING.
FT DNA_BIND 115 180 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 115 135 C4-TYPE.
FT ZN_FING 151 175 C4-TYPE.
FT DOMAIN 181 226 HINGE.
FT DOMAIN 227 446 LIGAND-BINDING.
FT VARSPLIC 1 87 MSTSHACPVAVRGHMTYPAAPYPLLPVIRGLSLPL
FT HGLHGHPPSPSCSPSPASVQACORTGGSGFAASTKWT
FT LKNA -> MFDCMDVLSVSGQILDFTYASPSCLMQEKA
FT LKACLSGFTQAEWHRRHTAQS (IN ISOFORM
FT BETA-2).
FT VARSPLIC 1 83 MISSING (IN ISOFORM BETA-4).
FT VARSPLIC 84 86 LNA -> MEN (IN ISOFORM BETA-4).
FT VARSPLIC 60 60 S -> T (IN ISOFORM BETA-1).
FT VARSPLIC 61 87 MISSING (IN ISOFORM BETA-1).
SQ SEQUENCE 482 AA; 53331 MW; 48E78E6C7D012515 CRC64;
Query Match 72.0%; Score 1753.5; DB 1; Length 482;
Best Local Similarity 76.1%; Pred. No. 8.4e-122;
```

```
Matches 338; Conservative 31; Mismatches 46; Indels 29; Gaps 3;
QY 33 GLPAPGYSPT-----LRTPLWNGS-NHSETQSSSEIEIVPSPSP 73
DB 46 GHPPSGCSTPSPASVQACQRTTGGSOFAASTKWTSPSLNAIETQSTSEELVPSPSP 105
QY 74 PPLPRIYKPCFVQDQSSGHHYGSACGCKGFFRRSIQKNMYTCHRDKNCIINKYTRN 133
DB 106 LPPRYIKPCFVQDQSSGHHYGSACGCKGFFRRSIQKNMYTCHRDKNCIINKYTRN 165
QY 134 PCOYCRLOKCFEYGMKSEYRNRNKKKVPKPCSESYTLTPEVGELEKVRKAHOET 193
DB 166 RCOYCRLOKCFEYGMKSEYRNRNKKKVPKPCSESYTLTPEVGELEKVRKAHOET 225
QY 194 FPALCOLGKYTTNNSRQYSLDLDLWDFSELSTKCIIVKVEFAKQLPGFTTLTIADQI 253
DB 226 FPSLCQLGKYTTNNSADHRVRLDGLWDFSELATKCIIVKVEFAKQLPGFTTLTIADQI 285
QY 254 TLKAAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFPLDLVFAFANQLLP 313
DB 286 TLKAAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFPLDLVFAFANQLLP 345
QY 314 LEMDDAETGLLSAICLCGRQDLEQDPRVDMLOEPLEALKVYVRRKRPSPHMFPKML 373
DB 346 LEMDDTETGLLSAICLCGRQDLEQDPRVDMLOEPLEALKVYVRRKRPSPHMFPKML 405
QY 374 MKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEGLDPLSGPGGGGDRGGSLP 433
DB 406 MKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEGLDPLSGPGGGGDRGGSLP 461
QY 434 PPPGSCSPSLSPSSNRSSPATHSP 457
DB 462 -----SPSVSPSSVSGVSQSP 479

RESULT 9
ID RRB_CORJA STANDARD; PRT; 455 AA.
AC Q9W6B3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retinoic acid receptor beta (RAR-beta).
GN RARB OR NR1B2.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN SEQUENCE FROM N.A.
RP Fu Z.W., Kato H., Sugahara K., Kubo T.;
RT "Molecular cloning of two isoforms of Japanese quail RAR beta and
RT their expression profile during embryogenesis and in the developing
RT tissues.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE
CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; BETA-1 (SHOWN HERE) AND BETA-2;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR1 SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF110729; AAD23397.1; -
CC HSP; P10826; 1HRA.
CC InterPro; IPR000536; Hormone_rec_lig.
CC InterPro; IPR001628; Znf_C4steroid.
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00047; STROIDFINGER.
CC PRODOM; PD000035; Znf_C4steroid; 1.
CC SMART; SM00430; HOL1; 1.
CC SMART; SM00399; Znf_C4; 1.
CC PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family; Alternative splicing.
FT DOMAIN 1 87 MODULATING (BY SIMILARITY).
FT DNA_BIND 88 153 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 88 108 C4-TYPE.
FT ZN_FING 124 148 C4-TYPE.
FT DOMAIN 154 199 HINGE (BY SIMILARITY).
FT DOMAIN 200 419 LIGAND-BINDING (BY SIMILARITY).
SQ SEQUENCE 455 AA; 50694 MW; 608951B106C4D41 CRC64;
Query Match 71.2%; Score 1736; DB 1; Length 455;
Best Local Similarity 77.3%; Pred. No. 1.5e-120;
Matches 331; Conservative 32; Mismatches 41; Indels 24; Gaps 2;
QY 30 PEKGLPAPGYSPTPLRTPLWNGSNHSIETQSSSEIEIVPSPSPPLPRIYKPCFVQDQK 89
DB 49 PTCGCTSP-----ASVETQSTSEELVPSPSPPLPRVYKPCFVQDQK 94
QY 90 SSGYHYGSACGCKGFFRRSIQKNMYTCHRDKNCIINKYVTRNPOCYCRLOKCFEYGM 149
DB 95 SSGYHYGSACGCKGFFRRSIQKNMYTCHRDKNCIINKYVTRNPOCYCRLOKCFEYGM 154
QY 150 KESVENDRNKKKVPKPCSESYTLTPVGVGELIERKRAHQETFPALCOLGKYTTNNS 209
DB 155 KESVENDRNKKKVPKPCSESYTLTPVGVGELIERKRAHQETFPALCOLGKYTTNNS 214
QY 210 EQRVSLDLDLWDFSELSTKCIIVKVEFAKQLPGFTTLTIADQITLLKAAACLDLILIRIC 269
DB 215 DHRVRLDGLWDFSELATKCIIVKVEFAKQLPGFTTLTIADQITLLKAAACLDLILIRIC 274
QY 270 TRYTPEDQMTFSDGLTLNRTQMHNAGFPLDLVFAFANQLPLEMDDAETGLLSAICL 329
DB 275 TRYTPEDQMTFSDGLTLNRTQMHNAGFPLDLVFAFANQLPLEMDDAETGLLSAICL 334
QY 330 ICGDRQDLEQDPRVDMLOEPLEALKVYVRRKRPSPHMFPMKMLITDLRSISAKGAER 389
DB 335 ICGDRQDLEQDPRVDMLOEPLEALKVYVRRKRPSPHMFPMKMLITDLRSISAKGAER 394
QY 390 VITLTKMEIPGSMPLIQEMLENSEGLDPLSGPGGGGDRGGSLPPPPGSCSPSLSPSSNR 449
DB 395 VITLTKMEIPGSMPLIQEMLENSEGLDPLSGPGGGGDRGGSLPPPPGSCSPSLSPSSNR 444
QY 450 SSPATHSP 457
DB 445 NSSVSQSP 452
RESULT 10
ID RRB_CHICK STANDARD; PRT; 455 AA.
AC P22448; Q90598; P27537; Q91354;
DT 01-AUG-1991 (Rel. 19, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Retinoic acid receptor beta (RAR-beta).
GN RARB OR NR1B2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
```


- |- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
- |- SUBCELLULAR LOCATION: Nuclear.
- |- ALTERNATIVE PRODUCTS: 2 isoforms; RAR-gamma1 (shown here) and RAR-gamma2 (AC P22932); are produced by alternative splicing. They differ only in their N-terminal regions.
- |- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN, A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
- |- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.

NR1 SUBFAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

EMBL; M24857; AAA53692.1; -
EMBL; M38258; AAA60254.1; -
EMBL; M57707; AAA63254.1; -
PIR; A33903; A33903.
PIR; A35573; A35573.
PIR; S26848; S26848.
PDB; 3LBD; O2-MAR-99.
PDB; 4LBD; O2-MAR-99.
TRANSFAC; T00720; -.
TRANSFAC; T01330; -.
Genew; HGNC:9866; RARG.
MIN; 180190; -.
InterPro; IPR000536; Hormone_rec_lig.
InterPro; IPR001723; Stdhrmn_receptor.
InterPro; IPR001628; ZnF_C4steroid.
Pfam; PF00104; hormone_rec; 1.
Pfam; PF00105; zf-C4; 1.
PRINTS; PR00398; STRDHORMONER.
PRINTS; PR00047; STROIDFINGER.
ProDom; PD000035; ZnF_C4steroid; 1.
SMART; SM00430; HOL1; 1.
SMART; SM00399; ZnF_C4; 1.
ProSite; PS00031; NUCLEAR_RECEPTOR; 1.
ProSite; Transcription regulation; DNA-binding; Nuclear protein;
Zinc-finger; Multigene family; Alternative splicing; 3D-structure.
DOMAIN 1 89
DNA_BIND 90 155 NUCLEAR RECEPTOR-TYPE.
ZN_FING 90 110 C4-TYPE.
ZN_FING 126 150 C4-TYPE.
DOMAIN 156 201 HINGE.
DOMAIN 202 421 LIGAND-BINDING.
SEQUENCE 454 AA; 50341 MW; 1EE27B2772DAFD CRG64;

```

Query Match	68.2%	Score 1662;	DB 1;	Length 454;
Best Local Similarity	72.8%;	Pred. No. 4.2e-115;		
Matches 319; Conservative 39; Mismatches 48; Indels 32; Gaps 4;				
QY	30	PEKGLPAG-PYSTP-----LRTPLWNG-----	SNHSIETQSSSSSEIYP	68
DB	16	PGSGYPCAGFFAAPPALRGSPPFEMLSPSFRGLGQDLPKEMASLSVETQTSSEMP	75	
QY	69	SPSPPLPLRIYPCFCVQDKSGSYHYGVSACGGCKGFFRRSTQKNMVTYCHRDKNCIIN	128	
DB	76	SSSPSPPPRRYKPCFVNCNDKSGSYHYGVSSCGCKGFFRRSTQKNMVTYCHRDKNCIIN	135	
QY	129	KVTRNPCYRLQCKCEVNGSKSVSRNDRNKKKVEYKPKCESSTYLTPEVGEGLIEKVKR	188	
DB	136	KVTRNRCYRLQCKCEVNGSKSVAVRNDRNKKKVEYKEGSPDSYELSPQLEELITYKSK	195	
QY	189	AHQETPALCOLGKYTTNNSSEORVSLDLDWDKFSLESTKCIIVKVFAPKQLPGFTTIT	248	
DB	196	AHQETPSCQLGKYTTNSADHRVGLDGLWDKFSLEATKCIIRIVEAPKRLPGFTGLS	255	

QY	249	IADQITLLKAAACDLILIRICTRYTPQDTWTFTSDGLTLNRTQMNAGEGPLTDLVF	308
Db	256	IADQITLLKAAACDLILIRICTRYTPQDTWTFTSDGLTLNRTQMNAGEGPLTDLVF	315
QY	309	NOLLPLEMDDAETGLLSAICLICGDRDLQPDQVDMQLQEPLEALKVYVRRPGRPHM	368
Db	316	GQLLPLEMDDTETGLLSAICLICGDRMDLEPEPKVQKLEPLAEURLTARRRPSQPM	375
QY	369	FPKMLMKITDLRSISAKGAERVITLKMIEPGSMPLPQIOMEENSEGLDTLSGPG	423
Db	376	FPRLMKITDLRGISTGGAERAITLKMIEPGMPPLIREMLENPEMFEDDSQGP	435
QY	424	-----GGRGDGGGLPPP	435
Db	436	SSEDEVFGGKGGLKSP	453
		STANDARD; PRT; 443 AA.	
RESULT	13		
RRG2	HUMAN		
ID	RCG2_HUMAN	STANDARD;	PRT; 443 AA.
IC	P22932;		
DT	01-AUG-1991 (Rel. 19, Created)		
DT	01-AUG-1991 (Rel. 19, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Retinoic acid receptor gamma-2 (RAR-gamma-2).		
GN	RARG OR NR1B3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE OF 1-63 FROM N.A.		
RP	MEDLINE=90207264; PubMed=2157210;		
RA	Kastner P., Krust A., Mendelsohn C., Garnier J.M., Zelent A.,		
RA	Leroy P., Staub A., Chambon P.;		
RT	"Murine isoforms of retinoic acid receptor gamma with specific		
RT	patterns of expression."		
RL	Proc. Natl. Acad. Sci. U.S.A. 87:2700-2704(1990).		

[2]
RP SEQUENCE OF 51-443 FROM N.A.
RX MEDLINE=89315787; Pubmed=2546152;
RA Kust A., Kastner P., Petkovich M., Zelent A., Chambon P.;
RT "A third human retinoic acid receptor, hRAR-gamma.,";
RL Proc. Natl. Acad. Sci. U.S.A. 86:5310-5314(1989).
[3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 167-412.
RX MEDLINE=96107307; Pubmed=7501014;
RA Renaud J.-P., Rochel N., Ruff M., Vivat V., Chambon P., Gronemeyer H.
RL Moras D.;
RT "Crystal structure of the RAR-gamma ligand-binding domain bound to
all-trans retinoic acid.,";
RL Nature 378:681-689(1995).
[4]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 167-412.
RX MEDLINE=98162555; Pubmed=9501913;
RA Klaholz B.P., Renaud J.-P., Mitschler A., Zusi C., Chambon P.,
RL Gronemeyer H., Moras D.;
RT "Conformational adaptation of agonists to the human nuclear receptor
RAR gamma.,";
RL Nat. Struct. Biol. 5:199-202(1998).
CC -1- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE
CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; RAR-gamma1 (AC P13631) and RAR-
CC gamma2 (shown here); are produced by alternative splicing. They
CC differ only in their N-terminal regions.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR1 SUBFAMILY.

```

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M24857; ; NOT_ANNOTATED_CDS.
EMBL; M32074; AAA60253.1; ;
PDB; 3LBD; 02-MAR-99.
PDB; 4LBD; 02-MAR-99.
DR TRANSFAC; T00720; ;
DR Genew; HGNC:9866; RARG.
DR MIM; 180190; ;
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrm_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR PRODom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family; Alternative splicing; 3D-structure.
FT DOMAIN 1 78
FT DNA_BIND 79 144 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 79 99 C4-TYPE.
FT ZN_FING 115 139 C4-TYPE.
FT DOMAIN 145 190 HINGE.
FT DOMAIN 191 410 LIGAND-BINDING.
SQ SEQUENCE 443 AA; 49307 MW; 4D709194F5111E86 CRC64;

Query Match 67.9%; Score 1655.5; DB 1; Length 443;
Best Local Similarity 71.6%; Pred. NO. 1.2e-114;
Matches 323; Conservative 35; Mismatches 68; Indels 25; Gaps 4;

QY 1 MYESVEVGPTNPFLVVDYFNQNRACLLPEKGL---PARGPYSTPLRPLW--NGSNHS 55
DB 1 MYDCHETAPGPRRLY-----GAGPGAGLLRRATGGSCFAGLESFAPQPASLOS 51
QY 56 IETQSSSEIEVPSPPPPLPRYKPCFVCQDKSSGYHYGSACEGCKGFFRRSIQKNM 115
DB 52 VETQSTSEEMVSPSPPPPPRVYKPCFVCNDXSSGYHYGSCEGCKGFFRRSIQKNM 111
QY 116 VYTCRDKNCLINKYRNPQCQCRLOKCFEVMGSKESVNRNDKKNKKVKEKPECSYTL 175
DB 112 VYTCRDKNCLINKYRNPQCQCRLOKCFEVMGSKESVNRNDKKNKKVKEKPECSYTL 171
QY 176 TPEVCELTEKYKHAQETFPALCOLGKGYTTNNSPQSVSLDLDLWCKSELSTKCIITV 235
DB 172 SPQLEELTKYKHAQETFPALCOLGKGYTTNNSPQSVSLDLDLWCKSELSTKCIITV 231
QY 236 EFAKQLPGFTTLTADQITLLKAACLDLILRICTRYPTPEQDTMTFSGLTLNRTQMHNA 295
DB 232 EFAKRLPGFTGLSTADQITLLKAACLDLILRICTRYPTPEQDTMTFSGLTLNRTQMHNA 291
QY 296 GFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGQRDLQEPDRVMDLQEPLEALK 355
DB 292 GFGPLTDLVFAFAGQLPLEMDDETGLLSAICLCGQRDLQEPDRVMDLQEPLEALK 351
QY 356 VYVKKRPSRPHMPKMLKTTDLRSISAKGNRVITLKMELPGSMPLIQEMLNSELG 415
DB 352 LYARRRPSQYMPFRLMLKTTDLRLGISTKGAERAITLKMELPGMPPLIREMLNPEMF 411
QY 416 DTLSCQPG-----GGRDGGGLPPP 435
DB 412 EDDSSQCPGPHNASEDEVPGGKGGLKSP 442

```

```

RESULT 14
RGL_MOUSE STANDARD; PRT; 458 AA.
ID RGL_MOUSE AC P18911;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Retinoic acid receptor gamma-A (RAR-gamma-A).
GN RARG OR NR1B3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89295563; PubMed=2544807;
RA Zelent A., Krust A., Petkovich M., Kastner P., Chambon P.;
RT "Cloning of murine alpha and beta retinoic acid receptors and a novel
RT receptor gamma predominantly expressed in skin.";
RL Nature 339:714-717(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90220619; PubMed=2157970;
RA Giguere V., Shago M., Zirngibl R., Tate P., Rossant J., Varmuza S.;
RT "Identification of a novel isoform of the retinoic acid receptor
RT gamma expressed in the mouse embryo.";
RL Mol. Cell. Biol. 10:2335-2340(1990).
RN [3]
RP SEQUENCE OF 1-74 FROM N.A.
RX MEDLINE=90207264; PubMed=2157210;
RA Kastner P., Krust A., Mendelsohn C., Garnier J.M., Zelent A.,
RA Leroy P., Staub A., Chambon P.;
RT "Murine isoforms of retinoic acid receptor gamma with specific
RT patterns of expression.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2700-2704(1990).
CC -!- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE
CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; RAR-GAMMA-A (SHOWN HERE) AND
CC RAR-GAMMA-B (AC P20787); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR1 SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X15848; CAA33845.1; ;
DR EMBL; M34476; AAA40035.1; ;
DR EMBL; M32068; AAA40032.1; ;
DR PIR; S05052; S05052.
DR PIR; A34714; A34714.
DR PIR; A35991; A35991.
DR HSSP; P22932; 3LBD.
DR TRANSFAC; T00717; ;
DR TRANSFAC; T01340; ;
DR MGD; MGI:97858; Rarg.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrm_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.

```



```
Db 228 IKIVEFAKRLPGFTGLSIADQITLLKAACLDILMLRICTRYTPQDWTFTSDGLTLNRQ 287
Qy 292 MHNAGFGPLTDLVFAFANOLLPLEMDDAETGLLSAICLICGDRQDLEQPDVDMLOEPLL 351
Db 288 MHNAGFGPLTDLVFAFAGQLLPLEMDDTETGLLSAICLICGDRMDLEEPEKYDKLOEPLL 347
Qy 352 EALKVYVRRRPSRPHMFKMLKITDLRSISAKGAERVITLKMEIPGSMPLIOEMLEN 411
Db 348 EALRLYARRDPAPKPYMPFRMLKITDLRGISTKGAERAITLKMEIPGMPPLIREMLN 407
Qy 412 SEGDLTSGQPG-----GGGRDGGGLPPPP 436
Db 408 PEMFEDDSKPGPHPKASSEDEAPGGGKRGQSQP 443
```

Search completed: March 29, 2003, 06:39:39
Job time : 24 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2003, 06:37:53 ; Search time 35 seconds
(without alignments)
1255.241 Million cell updates/sec

Title: US-09-691-220-2

Perfect score: 2437

Sequence: 1 MYSEVEGGTPNPFLVDF.....SCSPSLSPSSNRSSPAPHSP 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2127	87.3	452	1 A29491	retinoic acid rece
2	2124.5	87.2	452	2 S78481	retinoic acid rece
3	2119	87.0	955	4 C40045	probable transcrip
4	2099	86.1	453	2 I50674	retinoic acid rece
5	2089	85.7	462	2 S05050	retinoic acid rece
6	2058.5	84.5	464	2 A56558	retinoic acid rece
7	2026.5	83.2	448	2 B56558	retinoic acid rece
8	2008.5	82.4	444	2 I51256	retinoic acid rece
9	1972	80.9	458	2 S06123	retinoic acid rece
10	1926.5	79.1	459	2 A41977	retinoic acid rece
11	1779	73.0	448	2 S02827	retinoic acid rece
12	1771	72.7	448	2 S05051	retinoic acid rece
13	1757	72.1	448	2 A43786	retinoic acid rece
14	1733	71.1	455	2 S13512	retinoic acid rece
15	1663.5	68.3	499	2 I51257	retinoic acid rece
16	1662	68.2	454	1 A33903	retinoic acid rece
17	1655.5	67.9	443	1 C35991	retinoic acid rece
18	1651	67.7	458	2 A34714	retinoic acid rece
19	1633.5	67.0	447	2 B34714	retinoic acid rece
20	1627.5	66.8	442	2 A38592	retinoic acid rece
21	1621	66.5	454	2 S06124	retinoic acid rece
22	1620.5	66.5	476	2 B41977	retinoic acid rece
23	724	29.7	158	2 S02758	retinoic acid rece
24	658	27.0	410	2 S09178	retinoic acid rece
25	657.5	27.0	418	1 TVXLTA	thyroid hormone re
26	653	26.8	410	2 S14416	thyroid hormone re
27	653	26.8	418	2 B36067	thyroid hormone re
28	652	26.8	410	1 QRWSA1	thyroid hormone re
29	652	26.8	410	2 A40917	thyroid hormone re

30	646.5	26.5	391	2	I51097	thyroid hormone re
31	646	26.5	416	2	I51096	thyroid hormone re
32	645.5	26.5	408	1	TVCHVR	thyroid hormone re
33	639	26.2	373	2	I51165	gene c-erba-beta p
34	638	26.2	410	2	S06410	thyroid hormone re
35	630	25.9	373	2	C36067	thyroid hormone re
36	625.5	25.7	475	2	A57035	thyroid hormone be
37	625	25.6	373	2	D36067	thyroid hormone re
38	619.5	25.4	456	1	TVHUAR	thyroid hormone re
39	618	25.4	369	1	TVCHTB	thyroid hormone re
40	618	25.4	369	2	S58211	beta-thyroid hormo
41	615	25.2	461	2	A40377	thyroid hormone re
42	615	25.2	461	2	A31820	thyroid hormone re
43	614.5	25.2	398	1	TVFVVR	transforming prote
44	612.5	25.1	406	2	A44776	transforming prote
45	593.5	24.4	413	2	S14418	thyroid hormone re

ALIGNMENTS

RESULT 1

A29491

retinoic acid receptor alpha - human

C:Species: Homo sapiens (man)

C:Date: 04-Jun-1999 #sequence_revision 04-Jun-1999 #text_change 16-Jun-2000

C:Accession: A29491; S07272; S15598; A58975; A29740; S15599

R:Giguere, V.; Ong, E.S.; Segui, P.; Evans, R.M.

Nature 330, 624-629, 1987

A:Title: Identification of a receptor for the morphogen retinoic acid.

A:Reference number: A29491; MUID:88065922; PMID:2825036

A:Accession: A29491

A:Molecule type: mRNA

A:Residues: 1-462 <GIG>

A:Cross-references: GB:X06614; NID:g36156; PIDN:CAA29829.1; PID:g36157

R:Chambon, P. A human retinoic acid receptor which belongs to the family of nuclear recept

submitted to the EMBL Data Library, December 1988

A:Reference number: S07272

A:Accession: S07272

A:Molecule type: mRNA

A:Residues: 31-462 <CHA>

A:Cross-references: EMBL:X06538; NID:g35873; PIDN:CAA29787.1; PID:g35874

R:Petkovich, M.; Brand, N.J.; Krust, A.; Chambon, P.

Nature 330, 444-450, 1987

A:Title: A human retinoic acid receptor which belongs to the family of nuclear recept

A:Reference number: S06224; MUID:88065872; PMID:2825025

A:Accession: S06224

A:Molecule type: mRNA

A:Residues: 31-179, 'V', 181-462 <PET>

A:Cross-references: EMBL:X06538

A:Note: This sequence has been revised in reference S07272

A:Note: the authors translated the codon AAC for residue 360 as Lys; the sequence sho

R:Brand, N.J.; Petkovich, M.; Chambon, P.

Nucleic Acids Res. 18, 6799-6806, 1990

A:Title: Characterization of a functional promoter for the human retinoic acid recept

A:Reference number: S15594; MUID:91088249; PMID:2175878

A:Accession: S15598

A:Molecule type: DNA

A:Residues: 1-80 <BRAL>

A:Cross-references: EMBL:X56058; NID:g35876; PIDN:CAA39533.1; PID:g825712

A:Accession: A58975

A:Molecule type: DNA

A:Residues: 61-80 <BRA2>

A:Cross-references: GB:X58685; NID:g35878; PIDN:CAA41532.1; PID:gl335286

C:Genetics:

A:Gene: GDB:RARA

A:Cross-references: GDB:I20337; OMIM:I80240

A:Map position: 17q12-17q12

A:Introns: 60/1

A:Note: the first intron occurs before the initiator codon; the list of introns is in

C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology.

C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger

F:86-337/Domain: erba transforming protein homology <ERBA>

F:88-108/Region: zinc finger
F:124-148/Region: zinc finger

Query Match 87.3%; Score 2127; DB 1; Length 462;
Best Local Similarity 91.6%; Pred. No. 1.2e-148;
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;

QY 7 VGGTPNPFLVDFYNNRACLLPEKGLPAGPYSTPLRTPLWNGSNHSIETQSSSEEI 66
DB 32 LGLSP-PGALTTIQQH-----LPVSG-YSTP-----SPATIETQSSSEEI 71
QY 67 VPSPSPPLPRIYKPCFCVQDKSSGYHYGSACEGCKGFFRRSIQKNMYTCHRDNCI 126
DB 72 VPSPSPPLPRIYKPCFCVQDKSSGYHYGSACEGCKGFFRRSIQKNMYTCHRDNCI 131
QY 127 INKVTNRNCPQYCRLOKCFEYVMSKESVYRNDNRNKKKVPKPECSYTLTPEVGELETKV 186
DB 132 INKVTNRNCPQYCRLOKCFEYVMSKESVYRNDNRNKKKVPKPECSYTLTPEVGELETKV 191
QY 187 RKAQETFPALCOLGKYTTNNSQVSLDLDLWDFSELSTKCIITVEFAKOLPGFTT 246
DB 192 RKAQETFPALCOLGKYTTNNSQVSLDLDLWDFSELSTKCIITVEFAKOLPGFTT 251
QY 247 LTADQITLLKAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTOMHNAAGFPLTLVFA 306
DB 252 LTADQITLLKAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTOMHNAAGFPLTLVFA 311
QY 307 FANOLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMQLQEPLEALKVYVRRSRP 366
DB 312 FANOLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMQLQEPLEALKVYVRRSRP 371
QY 367 HMFPPKMLKITDLRSISAKGAERVITLKMETPGSMPLIQEMLNSENGLDLSQPGGGG 426
DB 372 HMFPPKMLKITDLRSISAKGAERVITLKMETPGSMPLIQEMLNSENGLDLSQPGGGG 431
QY 427 RDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
DB 432 RDGGGLAPPPGSCSPSLSPSSNRSSPATHSP 462

RESULT 2
S78481
retinoic acid receptor alpha-2 - eastern newt
C:Species: Notoththalmus viridescens, Triturus viridescens (eastern newt)
C>Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 20-Aug-1999
C:Accession: S78481; S26656
R:Ragsdale, C.W.
submitted to the EMBL Data Library, August 1992
A:Reference number: S78481
A:Accession: S78481
A:Molecule type: mRNA
A:Residues: 1-452 <RAG>
A:Cross-references: EMBL:Z14254; NID:G64123; PIDN:CAA78621.1; PID:G64124
R:Ragsdale Jr., C.W.; Gates, P.B.; Brookes, J.P.
Nucleic Acids Res. 20, 5851, 1992
A:Title: Identification and expression pattern of a second isoform of the newt alpha ret
A:Reference number: S26656; MUID:93087213; PMID:1333589
A:Accession: S26656
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-52 <RAW>
A:Cross-references: EMBL:Z14254
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger
F:79-331/Domain: erba transforming protein homology <ERBA>
F:81-101/Region: zinc finger CCCC motif
F:117-141/Region: zinc finger CCCC motif

Query Match 87.2%; Score 2124.5; DB 2; Length 452;
Best Local Similarity 87.1%; Pred. No. 1.8e-148;
Matches 399; Conservative 29; Mismatches 23; Indels 7; Gaps 3;

QY 1 MYESVEVGTPNPFLVDFYNNRACLLPEKGLPAGPYSTPLRTPLWNGSNHSIETQ 60

DB 1 MYDSVEVS--SPSPYIMIDFYSONRACLMDKGLGHPVPFGSPINRPHWSSSHSIETQS 58
QY 61 SSSSEIVSPSPPLPRIYKPCFCVQDKSSGYHYGSACEGCKGFFRRSIQKNMYTCH 120
DB 59 TSSEIVSPSPPLPRIYKPCFCVQDKSSGYHYGSACEGCKGFFRRSIQKNMYTCH 118
QY 121 RDKNCIINKVTNRNCPQYCRLOKCFEYVMSKESVYRNDNRNKKK-EVPPKPECSYTYITPEV 179
DB 119 RDKTCIINKVTNRNCPQYCRLOKCFEYVMSKESVYRNDNRNKKKQKQECTESYIITPEV 178
QY 180 GELTEKVRKAHQETFPALCOLGKYTTNNSQVSLDLDLWDFSELSTKCIITVEFAK 239
DB 179 EDLVEKVRKAHQETFPALCOLGKYTTNNSQVSLDLDLWDFSELSTKCIITVEFAK 238
QY 240 QLPGFTTITADQITLLKAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTOMHNAAGF 299
DB 239 QLPGFTTITADQITLLKAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTOMHNAAGF 298
QY 300 LTDLVFAFANOLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMQLQEPLEALKVYV 359
DB 299 LTDLVFAFANOLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMQLQEPLEALKVYV 358
QY 360 KRRPSRPHMFPKMLKITDLRSISAKGAERVITLKMETPGSMPLIQEMLNSENGLDLS 419
DB 359 KRRPNKPHMFPKMLKITDLRSISAKGAERVITLKMETPGSMPLIQEMLNSENGLDLS 418
QY 420 GQPGGGRRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
DB 419 GQP-----PRASSLAPPPGSCSPSLSPSSNRSSPATHSP 452

RESULT 3
C40045

probable transcription factor PML/retinoic acid receptor alpha mutant fusion protein
C:Species: Homo sapiens (man)
C>Date: 05-Jun-1992 #sequence_revision 17-Nov-1995 #text_change 20-Apr-2000
C:Accession: C40045; B40044
R:de The, H.; Lavau, C.; Marchio, A.; Chomienne, C.; Degos, L.; Dejean, A.
Cell 66, 675-684, 1991
A:Title: The PML-RARalpha fusion mRNA generated by the t(15;17) translocation in acu
A:Reference number: A40045; MUID:91347369; PMID:1652369
A:Accession: C40045
A:Molecule type: mRNA
A:Residues: 1-955 <DGT>
A:Cross-references: GB:S50916; NID:G234248; PIDN:AAB19602.1; PID:G234249
R:Kakizuka, A.; Miller Jr., W.H.; Umesono, K.; Warrell Jr., R.P.; Frankel, S.R.; Muri
Cell 66, 663-674, 1991
A:Title: Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fu
A:Reference number: A40044; MUID:91347368; PMID:1652368
A:Accession: B40044
A:Molecule type: mRNA
A:Residues: 1-394,553-955 <KAK>
A:Cross-references: GB:M73779; NID:G190124; PIDN:AAA60126.1; PID:G190125
C:Comment: This sequence is the chimeric product of a translocation mutation.

C:Genetics:
A:Gene: PML/RAR
A:Map position: 15q22/17q12
C:Keywords: alternative splicing; fusion protein
F:57-227/Region: cysteine-rich
F:581-601/Region: zinc finger motif
F:617-641/Region: zinc finger motif
Query Match 87.0%; Score 2119; DB 4; Length 955;
Best Local Similarity 90.9%; Pred. No. 1.2e-147;
Matches 409; Conservative 7; Mismatches 12; Indels 22; Gaps 3;

QY 30 PEKGLPA-----PGPYSTPLRTPLW-NGSNH-----SIETQSSSEIV 67
DB 506 PEQPRPSTSKAVSPHLLDGPSPRSPVIGSEVFLPNSNHNHVASGAGAAIETQSSSEIV 565
QY 68 PSPSPSPPLPRIYKPCFCVQDKSSGYHYGSACEGCKGFFRRSIQKNMYTCHRDNCI 127

Db 566 PSPSPPLPRIYKPCFVCODKSSGYHYGSACEGCKGFFRRSIQKNMVTCHRDKNCI 625
QY 128 NKVTRNCOYCRLOKCFEYVGSKEVSRNDRNKKKEVPKPCSESYTLTPVEGELIEKYR 187
Db 626 NKVTRNCOYCRLOKCFEYVGSKEVSRNDRNKKKEVPKPCSESYTLTPVEGELIEKYR 685
QY 188 KAHQETPALCOLGKYTTNNSSORVSLDIDLWDFSELSKCIITKVEFAKQLPGFTTL 247
Db 686 KAHQETPALCOLGKYTTNNSSORVSLDIDLWDFSELSKCIITKVEFAKQLPGFTTL 745
QY 248 TIADQITLLKAACDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGPLTDLVFAF 307
Db 746 TIADQITLLKAACDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGPLTDLVFAF 805
QY 308 ANQLLPLEMDDAETGLLSAICLICGDRDLEOPDRVDMQLQPLLBALKVYVYKRPRSRPH 367
Db 806 ANQLLPLEMDDAETGLLSAICLICGDRDLEOPDRVDMQLQPLLBALKVYVYKRPRSRPH 865
QY 368 MFPKMLMKITDLRSISAKGAERVITLKEIPGSMPLLIQEMLENSEGLDLSGQPGGGGR 427
Db 866 MFPKMLMKITDLRSISAKGAERVITLKEIPGSMPLLIQEMLENSEGLDLSGQPGGGGR 925
QY 428 DGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
Db 926 DGGGLAPPGSCSPSLSPSSNRSSPATHSP 955
RESULT 4
I50674
retinoic acid receptor alpha isoform 2-1 - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 20-Aug-1999
C:Accession: I50674
R:Michaille, J.J.; Kanzler, B.; Blanchet, S.; Garnier, J.M.; Dhoulailly, D.
Int. J. Dev. Biol. 39, 587-596, 1995
A:Title: Characterization of cDNAs encoding two chick retinoic acid receptor alpha isoforms
A:Reference number: I50674; MUID:96192630; PMID:8619957
A:Accession: I50674
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-453 <MIC>
A:Cross-references: EMBL:X78335; NID:gl089845; PIDN:CAA55134.1; PID:gl089846
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology
C:Keywords: zinc finger
F:79-330/Domain: erba transforming protein homology <ERBA>
Query Match 86.1%; Score 2099; DB 2: Length 453;
Best Local Similarity 88.0%; Pred. No. 1.4e-146;
Matches 403; Conservative 13; Mismatches 36; Indels 6; Gaps 4;
QY 1 MYSEVEGG-PTNPFLVDFYNONRACLPEKGLPAPGPYSTPLRTPLWNGSNHSIETQ 59
Db 1 MFEAGAVAGLPPGCPPLRMDCCGRGCLLPQ--CPPP-PTAPRAPHPWASGRSVETQ 57
QY 60 SSSSEETVPSPPPPLPRIYKPCFVCODKSSGYHYGSACEGCKGFFRRSIQKNMVTTC 119
Db 58 STSSEETVPSPPPPLPRIYKPCFVCODKSSGYHYGSACEGCKGFFRRSIQKNMVTTC 117
QY 120 HRDKNCILNKVTRNCOYCRLOKCFEYVGSKEVSRNDRNKKKEVPKPCSESYTLTPVEV 179
Db 118 HRDKNCILNKVTRNCOYCRLOKCFEYVGSKEVSRNDRNKKKEVPKPCSESYTLTPVEV 177
QY 180 GELIEKVRKAHOEFPAFCALCOLGKYTTNNSSQVSLDIDLWDFSELSKCIITKVEFAK 239
Db 178 EELIEKVRKAHOEFPAFCALCOLGKYTTNNSSQVSLDIDLWDFSELSKCIITKVEFAK 237
QY 240 QLPGFTTLTADQITLLKAACDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGP 299
Db 238 QLPGFTTLTADQITLLKAACDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGP 297
QY 300 LTDLVAFANQLLPLEMDDAETGLLSAICLICGDRDLEOPDRVDMQLQPLLBALKVYVYKR 359
Db 300 LTDLVAFANQLLPLEMDDAETGLLSAICLICGDRDLEOPDRVDMQLQPLLBALKVYVYKR 359

Db 298 LTDLVAFANQLLPLEMDDAETGLLSAICLICGDRDLEOPDRVDMQLQPLLBALKVYVYKR 357
QY 360 KRRPSPRPHMFKMLMKITDLRSISAKGAERVITLKEIPGSMPLLIQEMLENSEGLDLS 419
Db 358 KRRPSPRPHMFKMLMKITDLRSISAKGAERVITLKEIPGSMPLLIQEMLENSEGLDLS 417
QY 420 GPPGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
Db 418 GPPGGPRT--GGGLPPPGSCSPSLSPSSNRSSPATHSP 453
RESULT 5
S05050
retinoic acid nuclear receptor isoform alpha 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 20-Aug-1999
C:Accession: S05050; S23797; S13599; I56594
R:Zelent, A.; Krust, A.; Petkovich, M.; Kastner, P.; Chambon, P.
Nature 339, 714-717, 1989
A:Title: Cloning of murine alpha and beta retinoic acid receptors and a novel receptor
A:Reference number: S05050; MUID:89295363; PMID:2544807
A:Accession: S05050
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-462 <ZEL>
R:Kruyt, F.A.E.; van der Veer, L.J.; Mader, S.; van den Brink, C.E.; Feiljen, A.; Jor
Differentiation 49, 27-37, 1992
A:Title: Retinoic acid resistance of the variant embryonal carcinoma cell line RAC65
A:Reference number: S23797; MUID:92324516; PMID:1320576
A:Accession: S23797
A:Molecule type: mRNA
A:Residues: 1-390, 'A' <KRU>
A:Cross-references: EMBL:X57528; NID:g53890; PIDN:CAA04749.1; PID:g53891
A:Experimental source: embryonic carcinoma cell line RAC65
A:Note: this is the entire sequence of a truncated retinoic acid receptor alpha proc
R:Leroy, P.; Krust, A.; Zelent, A.; Mendelsohn, C.; Garnier, J.M.; Kastner, P.; Dier
EMBO J. 10, 59-69, 1991
A:Title: Multiple isoforms of the mouse retinoic acid receptor alpha are generated b
A:Reference number: S13599; MUID:91114713; PMID:1846598
A:Accession: S13599
A:Molecule type: mRNA
A:Residues: 1-73 <LER>
A:Cross-references: EMBL:X56572; NID:g53892; PIDN:CAA39919.1; PID:g53893
R:Heiermann, R.; Rentrop, M.; Lang, E.; Maelicke, A.
J. Recept. Res. 13, 693-709, 1993
A:Title: Cloning of several genes coding for retinoic acid nuclear receptors in the
A:Reference number: I56594; MUID:93195858; PMID:8383767
A:Accession: I56594
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-462 <RES>
A:Cross-references: GB:S56656; NID:g298621; PIDN:AAB25783.1; PID:g298622
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology
C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation; zi
F:86-337/Domain: erba transforming protein homology <ERBA>
F:88-108/Region: zinc finger
F:124-148/Region: zinc finger
Query Match 85.7%; Score 2089; DB 2: Length 462;
Best Local Similarity 90.0%; Pred. No. 7.7e-146;
Matches 406; Conservative 8; Mismatches 17; Indels 20; Gaps 4;
QY 7 VGGPTNPFLVDFYNONRACLPEKGLPAPGPYSTPLRTPLWNGSNHSITETSSSEEI 66
Db 32 LGGLSP-FGALVSLQHQ-----LPVSG-YSTP-----SPATETUSSSEEI 71
QY 67 VPSPPSPPLPRIYKPCFVCODKSSGYHYGSACEGCKGFFRRSIQKNMVTCHRDKNCI 126
Db 72 VPSPPSPPLPRIYKPCFVCODKSSGYHYGSACEGCKGFFRRSIQKNMVTCHRDKNCI 131
QY 127 INKVTNRCQYCRLOKCFEYVGSKEVSRNDRNKKKEVPKPCSESYTLTPVEGELIEKY 186
Db 132 INKVTNRCQYCRLOKCFEYVGSKEVSRNDRNKKKEVPKPCSESYTLTPVEGELIEKY 191

QY 187 RKAHOETFPALCOLGKYTTNNSSEORVSLDIDLWDKFSLSLTKCIKTYVERAKQLPGFTT 246
Db 192 RKAHOETFPALCOLGKYTTNNSSEORVSLDIDLWDKFSLSLTKCIKTYVERAKQLPGFTT 251
QY 247 LTIADQITLLKAACLDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFPLTDLVFA 306
Db 252 LTIADQITLLKAACLDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFPLTDLVFA 311
QY 307 FANQLLPLEMDDAETGLLSAICLICGDRDLQEPDRVMDLQEPLEALKVYVRKRRSP 366
Db 312 FANQLLPLEMDDAETGLLSAICLICGDRDLQEPDRVMDLQEPLEALKVYVRKRRSP 371
QY 367 HMPFKMLMKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDITLSGOGGGG 426
Db 372 HMPFKMLMKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDITLSGOGGGG 431
QY 427 RDGGGLPPPGSCPSLSPSSNRSSPATHSP 457
Db 432 RDGGGLAPPGSCPSLSPSSNRSSPATHSP 462

RESULT 6

A56558

retinoic acid receptor isoform alpha 2.1 - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 31-Oct-1997

C:Accession: A56558

R:Sharpe, C.R.

Mech. Dev. 39, 81-93, 1992

A:Title: Two isoforms of retinoic acid receptor alpha expressed during Xenopus developme

A:Reference number: A56558; MUID:93144180; PMID:1336976

A:Accession: A56558

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-464 <SHA>

A:Experimental source: embryos

A:Note: sequence extracted from NCBI backbone (NCBIN:123865, NCBIPI:123867)

C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology

C:Keywords: zinc finger

F:91-342/Domain: erba transforming protein homology <ERBA>

Query Match 84.5%; Score 2058.5; DB 2; Length 464;

Best Local Similarity 85.6%; Pred. No. 1.4e-143;

Matches 393; Conservative 21; Mismatches 40; Indels 5; Gaps 2;

QY 1 MYSEVEVG--GPTNPPLVDFYNQNRACLLPEKGLPAGPYSTPLRPLNNGSNHSTET 58

Db 9 MYENVDPALASSPTRFHMDFYSHNRQCLLQEKIGITIHPIYGTPLRQHWSSNHSIET 68

QY 59 QSSSEIEIVPSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKNMYT 118

Db 69 QSTSEIEIVPSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKNMYT 128

QY 119 CHRDKNCIINKVTRNRCQYRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPE 178

Db 129 CHRDKNCIINKVTRNRCQYRLQKCFEVMGSKESVRNDRNKKKEVPKPEAIESYILSPE 188

QY 179 VGLIEKVRKAHOETFPALCOLGKYTTNNSSEORVSLDIDLWDKFSLSLTKCIKTYVEFA 238

Db 189 TDLIEKVRKAHOETFPALCOLGKYTTNNSSEORVSLDIDLWDKFSLSLTKCIKTYVEFA 248

QY 239 KQLPGFTTLTIAQITLLKAACLDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFG 298

Db 249 KQLPGFTTLTIAQITLLKAACLDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFG 308

QY 299 PLTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDLQEPDRVMDLQEPLEALKVYV 358

Db 309 PLTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDLQEPDRVMDLQEPLEALKVYV 368

QY 359 RKRPRSRPHMPKMLMKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDITL 418

Db 369 RTRRPQKPHMPKMLMKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDITL 428

QY 419 SQGPQGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457
Db 429 G---GGASSDAPVTVPAGSCSPSLSPSSSTHSSPSHSP 464

RESULT 7

B56558

retinoic acid receptor isoform alpha 2.2 - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 20-Aug-1999

C:Accession: B56558

R:Sharpe, C.R.

Mech. Dev. 39, 81-93, 1992

A:Title: Two isoforms of retinoic acid receptor alpha expressed during Xenopus devel

A:Reference number: A56558; MUID:93144180; PMID:1336976

A:Accession: B56558

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-448 <SHA>

A:Cross-references: GB:X87365; GB:S53907; GB:S53908; NID:9836654; PIDN:CAA60791.1;

A:Experimental source: embryos

A:Note: sequence extracted from NCBI backbone (NCBIN:123871, NCBIPI:123874)

C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology

C:Keywords: zinc finger

F:75-326/Domain: erba transforming protein homology <ERBA>

Query Match 83.2%; Score 2026.5; DB 2; Length 448;

Best Local Similarity 84.9%; Pred. No. 2.9e-141;

Matches 388; Conservative 22; Mismatches 38; Indels 9; Gaps 3;

QY 1 MYSEVEVGPTNPPLVDFYNQNRACLLPEKGLPAGPYSTPLRPLNNGSNHSTETOS 60

Db 1 MYENVDV---SPTHVMMDFYSHNRQCLWPEKRI---NPYGTPLGTQHWSSNHSIETOS 54

QY 61 SSSEIEIVPSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKNMYTCH 120

Db 55 TSSEIEIVPSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKNMYTCH 114

QY 121 RDKNCIINKVTRNRCQYRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPEVG 180

Db 115 RDKNCIINKVTRNRCQYRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPEVQ 174

QY 181 ELIEKVRKAHOETFPALCOLGKYTTNNSSEORVSLDIDLWDKFSLSLTKCIKTYVEFAKO 240

Db 175 DLIEKVRKAHOETFPALCOLGKYTTNNSSEORVSLDIDLWDKFSLSLTKCIKTYVEFAKO 234

QY 241 LPGFTTLTIAQITLLKAACLDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGPL 300

Db 235 LPGFTTLTIAQITLLKAACLDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGPL 294

QY 301 TDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDLQEPDRVMDLQEPLEALKVYVRK 360

Db 295 TDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDLQEPDRVMDLQEPLEALKVYVRN 354

QY 361 RPRSRPHMPKMLMKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDITLSG 420

Db 355 RPRSRPHMPKMLMKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDITLSG 413

QY 421 QPGGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457

Db 414 --GGASSDPTVTPVAPGSCPSLSPSSSTHSSPSHSP 448

RESULT 8

I51256

retinoic acid receptor alpha - zebra fish

C:Species: Brachydanio rerio (zebra fish)

C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 20-Aug-1999

C:Accession: I51256

R:Joore, J.; van der Lans, G.B.; Lanser, P.H.; Vervaaert, J.M.; Zivkovic, D.; Speksni;

Mech. Dev. 46, 137-150, 1994

A:Title: Effects of retinoic acid on the expression of retinoic acid receptors during

A:Reference number: I51256; MUID:95001557; PMID:7918098
A:Accession: I51256
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-444 <DOO>
A:Cross-references: GB:S74155; NID:g704369; PIDN:AA32276.1; PID:g704370
C:Genetics:
A:Gene: zRARalpha
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology
F:73-324/Domain: erba transforming protein homology <ERBA>

Query Match 82.4%; Score 2008.5; DB 2; Length 444;
Best Local Similarity 82.9%; Pred. No. 6e-140;
Matches 379; Conservative 35; Mismatches 30; Indels 13; Gaps 3;

QY 1 MYSEVEGGTPNPLVDFVFNQNRACLLPEKGLPAPGYPSTPLRPLWNGSNHSIETQS 60
DB 1 MYSEVDV-----NPLMDYNYNQSGCLIPDK---MPHPFSSIRHQHWSGNSHSIETQS 52
QY 61 SSSSEIVPSPSPPLPRIYKPCFCQDKSSGYHYGVSAECGCKGFFRRSIQKNMVTYCH 120
DB 53 TSSEIVPSPSPSPPLPRIYKPCFCQDKSSGYHYGVSAECGCKGFFRRSIQKNMVTYCH 112
QY 121 RDKNCIINKVTRNRCQYCRQKCEVGMKESVNRNDRNKKKEVPKPCSESYTLTPEVG 180
DB 113 REKNCIINKVTRNRCQYCRQKCEVGMKESVNRNDRNKKKEEKPECTENYTLSPDTE 172
QY 181 ELIEKVRKAHQETFPALCOLGKYTTNNSSEQVSLDIDLWKFSELSTKCIITVEFAKQ 240
DB 173 QMIDRVKKAHQETFPALCOLGKYTTNNSSEQVSLDIDLWKFSELSTKCIITVEFAKQ 232
QY 241 LPGFTTLTIADQITLLKACDLILILICTRYTPQDTMTFSDGLTLNRTQMHNAGFGPL 300
DB 233 LPGFTTLTIADQITLLKACDLILILICTRYTPQDTMTFSDGLTLNRTQMHNAGFGPL 292
QY 301 TDLVAFANQLLPLEMDAETGLLSAICLCGRDRLDLPDQVDMLOEPLLEALKYVVRK 360
DB 293 TDLVAFANQLLPLEMDAETGLLSAICLCGRDRLDLPDQVDMLOEPLLEALKYVVRN 352
QY 361 RRPGRPHFPKMLKITDLRSISAKGAERTVTLKMEIPGSMPLIOEMLENSEGLDLSG 420
DB 353 RRPKHPMFPMKMLKITDLRSISAKGAERTVTLKMEIPGSMPLIOEMLENSEGLSSG 412
QY 421 QPGGGRDGLLPPPGSCSPSLSPSSNRSSPATHSP 457
DB 413 AQGSRAS-----ATTPGSCSPSLSPNSAQSPPQTQS 444

RESULT 9
S06123
retinoic acid receptor alpha - eastern newt
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 20-Aug-1999
C:Accession: S06123
R:Ragsdale Jr., C.W.; Petkovich, M.; Gates, P.B.; Chambon, P.; Brockes, J.P.
Nature 341, 654-657, 1989
A:Title: Identification of a novel retinoic acid receptor in regenerative tissues of the
A:Reference number: S06123; MUID:90015190; PMID:2552324
A:Accession: S06123
A:Molecule type: mRNA
A:Residues: 1-458 <RAG>
A:Cross-references: EMBL:X17585; NID:g64125; PIDN:CAA35602.1; PID:g64126
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger
F:85-337/Domain: erba transforming protein homology <ERBA>
F:87-107/Region: zinc finger
F:123-147/Region: zinc finger

Query Match 80.9%; Score 1972; DB 2; Length 458;
Best Local Similarity 88.3%; Pred. No. 3e-137;
Matches 379; Conservative 16; Mismatches 18; Indels 16; Gaps 4;

QY 33 GLPAPGP---YSTPLRPLWNGSNHSIETQSSSEIVSPSPPLPRIYKPCFCVQDK 89

DB 42 GIPHPLVSAVSTP-----SPATIEQTSSSEIVSPSPPLPRIYKPCFCVQDK 93
QY 90 SSGYHYGVSAECGCKGFFRRSIQKNMVTYCHRDKNCIINKVTRNRCQYCRQKCEVGMK 149
DB 94 SSGYHYGVSAECGCKGFFRRSIQKNMVTYCHRDKNCIINKVTRNRCQYCRQKCEVGMK 153
QY 150 KESVRNDRNKKKK-EVPPKPCSESYTLTPEVGELIEKVRKAHQETFPALCOLGKYTTNS 208
DB 154 KESVRNDRNKKKKQKQAPKQECTESYIITPEVEDLVEKVRKAHQETFPALCOLGKYTTNS 213
QY 209 SEQVSLDIDLWKFSELSTKCIITVEFAKOLPGFTTLTIADQITLLKAAACDLILIRI 268
DB 214 SEERYSLDIDLWKFSELSTKCIITVEFAKOLPGFTTLTIADQITLLKAAACDLILIRI 273
QY 269 CRYTPPEODTMTFSDGLTLNRTQMHNAGFGPLTDLVAFANQLLPLEMDAETGLLSAIC 328
DB 274 CRYTPPDQDTMTFSDGLTLNRTQMHNAGFGPLTDLVAFANQLLPLEMDAETGLLSAIC 333
QY 329 LICGRDRLDLPDQVDMLOEPLLEALKYVVRKRRSPRPHFPKMLKITDLRSISAKGAE 388
DB 334 LICGRDRLDLPDQVDMLOEPLLEALKYVVRKRRSPRPHFPKMLKITDLRSISAKGAE 393
QY 389 RVITLKMETPGSMPLIOEMLENSEGLDLSGPGGGRDGLLPPPGSCSPSLSPSSN 448
DB 394 RVITLKMETPGSMPLIOEMLENSEGLDLSGPGGGRDGLLPPPGSCSPSLSPSSN 449
QY 449 RSSPATHSP 457
DB 450 RSSPTSHSP 458

RESULT 10
A41977
retinoic acid receptor alpha - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 07-Feb-1997
C:Accession: A41977
R:Blumberg, B.; Mangelsdorf, D.J.; Dyck, J.A.; Bittner, D.A.; Evans, R.M.; De Robertis
Proc. Natl. Acad. Sci. U.S.A. 89, 2321-2325, 1992
A:Title: Multiple retinoid-responsive receptors in a single cell: families of retinoid
A:Reference number: A41977; MUID:92196110; PMID:1312717
A:Accession: A41977
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-459 <BLU>
A:Experimental source: egg
A:Note: sequence extracted from NCBI backbone (NCBIP:88334)
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology
C:Keywords: zinc finger
F:86-337/Domain: erba transforming protein homology <ERBA>
F:88-108/Region: zinc finger
F:124-148/Region: zinc finger

Query Match 79.1%; Score 1926.5; DB 2; Length 459;
Best Local Similarity 81.6%; Pred. No. 6.7e-134;
Matches 377; Conservative 17; Mismatches 43; Indels 25; Gaps 4;

QY 10 PTPNPLVDFVFNQNRACLLPEK-----GLPAP-----GPYSTPLRPLWNGSNHS 55
DB 9 PPGPGHINGFHPHYAFFPHMLGMSPTGGLPGVQHPPLSGYSTP-----SPAT 60
QY 56 IETQSSSEIVSPSPPLPRIYKPCFCQDKSSGYHYGVSAECGCKGFFRRSIQKNM 115
DB 61 IETQSSSEIVSPSPPLPRIYKPCFCQDKSSGYHYGVSAECGCKGFFRRSIQKNM 120
QY 116 VYTCRDKNCIINKVTRNRCQYCRQKCEVGMKESVNRNDRNKKKEVPKPCSESYTL 175
DB 121 VYTCRDKNCIINKVTRNRCQYCRQKCEVGMKESVNRNDRNKKKEVPKPCSESYTL 180
QY 176 TPEVGELIEKVRKAHQETFPALCOLGKYTTNNSSEQVSLDIDLWKFSELSTKCIITKV 235
DB 181 SPETQDLIEKVRKAHQETFPALCOLGKYTTNNSSEQVSLDIDLWKFSELSTKCIITKV 240

QY 236 EFAKQLPGFTTITADQITLLKAACLDILILIRICTRYTPEDQTMFSDGLTLNRTQMNA 295
|||||
Db 241 EFAKQLPGFTTITADQITLLKAACLDILILIRICTRYTPEDQTMFSDGLTLNRTQMNA 300
QY 296 GFGPLTDLVFAFANOLLPLEMDDAETGLLSAICLCGRQDLQEPDRYDMLQEPLEALK 355
Db 301 GFGPLTDLVFAFANOLLPLEMDDAETGLLSAICLCGRQDLQEPDRYDMLQEPLEALK 360
QY 356 VYVKKRPSRPHMPKMLKMTDLRSISAKGAERVITLTKMEIPGSMPLLIQEMLENSEGL 415
|||||
Db 361 IYVTRRRPQKHPMKMLKMTDLRSISAKGAERVITLTKMEIPGSMPLLIQEMLENSEGL 420
QY 416 DTLGQPGGGGRDGGGLPPPPGSCPSLSPSSNRSSPATHSP 457
Db 421 DTLG---GGASSDAPVTPVAPGSCPSLSPSSSTHSSPATHSP 459

RESULT 11

S02827

retinoic acid receptor beta-2 - human

N:Alternate names: thyroid hormone receptor-related protein

C:Species: Homo sapiens (man)

C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 20-Aug-1999

C:Accession: S02827; A29492; S30489; I57458

R:Benbrook, D.; Lernhardt, E.; Pfahl, M.

Nature 333, 669-672, 1988

A:Title: A new retinoic acid receptor identified from a hepatocellular carcinoma.

A:Reference number: S02827; MUID:88232961; PMID:2836738

A:Accession: S02827

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-448 <BEN>

R:de The, H.; Marchio, A.; Tiollais, P.; Dejean, A.

Nature 330, 667-670, 1987

A:Title: A novel steroid thyroid hormone receptor-related gene inappropriately expressed

A:Reference number: A29492; MUID:88065931; PMID:2835037

A:Accession: A29492

A:Molecule type: mRNA

A:Residues: 1-406 'M', 408-448 <DE>

A:Cross-references: GB:Y00291

A:Experimental source: liver

R:Shen, S.; Kruyt, F.A.E.; den Hertog, J.; van der Saag, P.T.; Kruijjer, W.

DNA Seq. 2, 111-119, 1991

A:Title: Mouse and human retinoic acid receptor beta-2 promoters: sequence comparison an

A:Reference number: S30489; MUID:92135889; PMID:1663808

A:Accession: S30489

A:Molecule type: mRNA

A:Residues: 1-52 <SHE>

A:Cross-references: EMBL:X56849

R:Dejean, A.; de The, H.

Mol. Biol. Med. 7, 213-222, 1990

A:Title: Hepatitis B virus as an insertional mutagen in a human hepatocellular carcinoma

A:Reference number: I57458; MUID:91014673; PMID:2170809

A:Accession: I57458

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 54-102 <RES>

A:Cross-references: GB:M57445; NID:g184476; PIDN:AAA58728.1; PID:g184477

C:Genetics:

A:Gene: HUP

C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology

C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation; zinc f

F:79-330/Domain: erba transforming protein homology <ERBA>

F:81-101/Region: zinc finger

F:117-141/Region: zinc finger

Query Match 73.08; Score 1779; DB 2; Length 448;

Best Local Similarity 74.84; Pred. No. 4.5e-123;

Matches 344; Conservative 41; Mismatches 57; Indels 18; Gaps 5;

QY 1 MYESVEGGTTPNPLVVDVFNQN-RACLLPEKGLPAPGPYSTPLRPLWNGSN--HSIE 57

Db 1 MFDCMDVLSVSGQ--ILDFTASPSMLOEAKLA---CFLGTLQTEHQHHRHTAQISIE 55
QY 58 TQSSSEETVPPSPPPPLPRYKPCFVQODKSSGGHYGVSAEGCKGFFRRIQKNMY 117
Db 56 TQSTSEELVPPSPPLPPRVYKPCFVQODKSSGGHYGVSAEGCKGFFRRIQKNMY 115
QY 118 TCHRNKNCIINKVTRNPOCYCRLOKCFEYGMKSKEVNRNDRNKKKKEVPKPECSSESYTLTP 177
Db 116 TCHRNKNCIINKVTRNPOCYCRLOKCFEYGMKSKEVNRNDRNKKKKEVPECTESYEMTA 175
QY 178 EYGLIEKVRKAHQETFPALCOLGYTTNNSEQRVSLDIDLWDFSELSKCIITVEF 237
Db 176 ELDDLTEKIRKAHQETFPSCQLGKYTTNSSADHRVRLDGLWDFSELSKCIITVEF 235
QY 238 AKQLPGFTTITADQITLLKAACLDILILIRICTRYTPEDQTMFSDGLTLNRTQMNA 297
Db 236 AKRLPGFTTITADQITLLKAACLDILILIRICTRYTPEDQTMFSDGLTLNRTQMNA 295
QY 298 GPLTDLVFAFANOLLPLEMDDAETGLLSAICLCGRQDLQEPDRYDMLQEPLEALKVY 357
Db 296 GPLTDLVFAFANOLLPLEMDDTETGLLSAICLCGRQDLQEPDRYDMLQEPLEALKVY 355
QY 358 YKRRPSRPHMPKMLKMTDLRSISAKGAERVITLTKMEIPGSMPLLIQEMLENSEGLDT 417
Db 356 IRKRRPSRPHMPKMLKMTDLRSISAKGAERVITLTKMEIPGSMPLLIQEMLENSEGH 415
QY 418 LSGQPGGGGRDGGGLPPPPGSCPSLSPSSNRSSPATHSP 457
Db 416 LTPSSSGNTAEH-----SPSISPSSESVNSGVQSQP 445

RESULT 12

S05051

retinoic acid receptor beta-2 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 20-Aug-1999

C:Accession: S05051; S13516; S30490

R:Zelent, A.; Krust, A.; Petkovich, M.; Kastner, P.; Chambon, P.

Nature 339, 714-717, 1989

A:Title: Cloning of murine alpha and beta retinoic acid receptors and a novel recept

A:Reference number: S05050; MUID:89295563; PMID:2544807

A:Accession: S05051

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-448 <ZELL>

R:Zelent, A.; Mendelsohn, C.; Kastner, P.; Krust, A.; Garnier, J.M.; Ruffenach, F.;

EMBO J. 10, 71-81, 1991

A:Title: Differentially expressed isoforms of the mouse retinoic acid receptor beta

A:Reference number: S13515; MUID:91114714; PMID:1846599

A:Accession: S13516

A:Molecule type: mRNA

A:Residues: 1-59 <ZEL2>

A:Cross-references: EMBL:X56573; NID:g53903; PIDN:CAA39920.1; PID:g53904

R:Shen, S.; Kruyt, F.A.E.; den Hertog, J.; van der Saag, P.T.; Kruijjer, W.

DNA Seq. 2, 111-119, 1991

A:Title: Mouse and human retinoic acid receptor beta-2 promoters: sequence compariso

A:Reference number: S30489; MUID:92135889; PMID:1663808

A:Accession: S30490

A:Molecule type: DNA

A:Residues: 1-52 <SHE>

A:Cross-references: EMBL:X56850

C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology

C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation; zif

F:79-330/Domain: erba transforming protein homology <ERBA>

F:81-101/Region: zinc finger CCCC motif

F:117-141/Region: zinc finger CCCC motif

Query Match

Best Local Similarity 72.78; Score 1771; DB 2; Length 448;

Matches 344; Conservative 40; Mismatches 58; Indels 18; Gaps 5;

QY 1 MYESVEGGTTPNPLVVDVFNQN-RACLLPEKGLPAPGPYSTPLRPLWNGSN--HSIE 57

Db 1 MFDCMDVLSVSPGQ--ILDEVTSAPSSCMLQEKALKACLSGFTQAE--WQHRHTAQASIE 55
QY 58 TQSSSEIEVPSPPPLPRIYKPCVQDKSSGYHYGVSAACGCKGFFRRSIQKNMY 117
Db 56 TQSTSEELVPSPPPLPRVYKPCVQDKSSGYHYGVSAACGCKGFFRRSIQKNMY 115
QY 118 TCHRDKNCIINKNVTRNCPQYCRLOKCFEVMGSKESVNRDRNKKKPKCESYTLTP 177
Db 116 TCHRDKNVINKVTRNCPQYCRLOKCFEVMGSKESVNRDRNKKKPKCESYTLTP 175
QY 178 EVGELIEKVRKAHOETFPALCQKGYTTNNSSSEORVSLDLDLWDFSELSTKCIKIVF 237
Db 176 ELDDLTEKIRKAHOETFPSCQLGKYTTNSADHRVLDLGLWDFSELATKCIKIVF 235
QY 238 AKQLPGFTTLTIADQITLLKAACLDILILRICTRYTPQDTMTFSDGLTLNRTQMHNAGF 297
Db 236 AKRLPGFTGLTIADQITLLKAACLDILILRICTRYTPQDTMTFSDGLTLNRTQMHNAGF 295
QY 298 GPLTDLVAFANQLLPLEMDDAETGLLSAICLCIGDRQDLEOPDRVMDLQEPLEALKY 357
Db 296 GPLTDLVTFANQLLPLEMDDETETGLLSAICLCIGDRQDLEETPKVDKLOEPLEALKY 355
QY 358 VRKRSPRPHFPMKMLKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDT 417
Db 356 IRKRSPKPHFPMKMLKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDT 415
QY 418 LSGOPGGGRDGGGLPPPGSCSPSLSSNRSSPATHSP 457
Db 416 LTPSSSGNIAEH-----SPSVSPSSVNSGVQSOP 445

RESULT 13

retinoic acid receptor beta-4 - chicken
C:Species: Gallus gallus (chicken)
C:Date: 11-Dec-1992 #sequence_revision 30-Jan-1993 #text_change 20-Aug-1999
C:Accession: A43786; S14291; A60129; S17072; S16243
R:Smith, S.M.; Eichele, G.
A:Title: Temporal and regional differences in the expression pattern of distinct retinoic acid receptor beta-4 isoforms in the chicken embryo.
A:Reference number: A43786; MUID:91199956; PMID:1849811
A:Accession: A43786
A:Molecule type: mRNA
A:Residues: 1-448 <SMI>
A:Cross-references: GB:X59473; NID:963753; PIDN:CAA42077.1; PID:g63754
A:Note: The authors translated the codon ACT for residue 96 as Ser, TCC for residue 167
R:Noji, S.; Nohno, T.; Koyama, E.; Muto, K.; Ohya, K.; Ohsugi, K.
Nature 350, 83-86, 1991
A:Title: Retinoic acid induces polarizing activity but is unlikely to be a morphogen in the chicken embryo.
A:Reference number: S14291; MUID:91163640; PMID:1848357
A:Accession: S14291
A:Molecule type: mRNA
A:Residues: 1-95, 'S', '97-448 <NOJ>
A:Cross-references: EMBL:X57340; NID:g62976; PIDN:CAA40616.1; PID:g62977
R:Rowe, A.; Richman, J.M.; Brickell, P.M.
Development 111, 1007-1016, 1991
A:Title: Retinoic acid treatment alters the distribution of retinoic acid receptor-beta isoforms in the chicken embryo.
A:Reference number: A60129; MUID:91347912; PMID:1652423
A:Accession: A60129
A:Molecule type: mRNA
A:Residues: 23-24, 26-95, 'S', '97-165, 'D', '167-448 <ROW>
A:Cross-references: GB:S63196; NID:g234370; PIDN:AAB19628.1; PID:g234371
R:Nohno, T.
submitted to the EMBL Data Library, January 1991
A:Reference number: S17072
A:Accession: S17072
A:Molecule type: mRNA
A:Residues: 15-95, 'S', '97-434 <NOH>
A:Cross-references: EMBL:X57339
R:Nohno, T.; Muto, K.; Noji, S.; Saito, T.; Taniguchi, S.
Biochim. Biophys. Acta 1089, 273-275, 1991
A:Title: Isoforms of retinoic acid receptor beta expressed in the chicken embryo.

A:Reference number: S16243; MUID:91274365; PMID:1647216
A:Accession: S16243
A:Molecule type: mRNA
A:Residues: 15-60 <NOH2>
A:Cross-references: EMBL:X57339
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology
C:Keywords: alternative splicing; DNA binding; zinc finger
F:79-330/Domain: erba transforming protein homology <ERBA>
F:81-101/Region: zinc finger
F:117-141/Region: zinc finger

Query Match 72.1%; Score 1757; DB 2; Length 448;
Best Local Similarity 74.3%; Pred. No. 1.9e-121;
Matches 342; Conservative 42; Mismatches 58; Indels 18; Gaps 5;
QY 1 MYESVEVGPPNPPLVVDVFNQNRACLLPEKGLPAFPYSPYTLRPLW--NGSNHSIE 57
Db 1 MFDCMDVLSVSPGQ--ILDEVTSAPSSCMLQEKALKACLSGFTQAE--WQHRHTAQASIE 55
QY 58 TQSSSEIEVPSPPPLPRIYKPCVQDKSSGYHYGVSAACGCKGFFRRSIQKNMY 117
Db 56 TQSTSEELVPSPPPLPRVYKPCVQDKSSGYHYGVSAACGCKGFFRRSIQKNMY 115
QY 118 TCHRDKNCIINKNVTRNCPQYCRLOKCFEVMGSKESVNRDRNKKKPKCESYTLTP 177
Db 116 TCHRDKNVINKVTRNCPQYCRLOKCFEVMGSKESVNRDRNKKKPKCESYTLTP 175
QY 178 EVGELIEKVRKAHOETFPALCQKGYTTNNSSSEORVSLDLDLWDFSELSTKCIKIVF 237
Db 176 ELDDLTEKIRKAHOETFPSCQLGKYTTNSADHRVLDLGLWDFSELATKCIKIVF 235
QY 238 AKQLPGFTTLTIADQITLLKAACLDILILRICTRYTPQDTMTFSDGLTLNRTQMHNAGF 297
Db 236 AKRLPGFTGLTIADQITLLKAACLDILILRICTRYTPQDTMTFSDGLTLNRTQMHNAGF 295
QY 298 GPLTDLVAFANQLLPLEMDDAETGLLSAICLCIGDRQDLEOPDRVMDLQEPLEALKY 357
Db 296 GPLTDLVTFANQLLPLEMDDETETGLLSAICLCIGDRQDLEETPKVDKLOEPLEALKY 355
QY 358 VRKRSPRPHFPMKMLKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDT 417
Db 356 IRKRSPKPHFPMKMLKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDT 415
QY 418 LSGOPGGGRDGGGLPPPGSCSPSLSSNRSSPATHSP 457
Db 416 LTPSSSGNIAEH-----SPSVSPSSVNSGVQSOP 445

RESULT 14

S13512
retinoic acid receptor beta-1 - chicken
C:Species: Gallus gallus (chicken)
C:Date: 21-Nov-1993 #sequence_revision 22-Apr-1995 #text_change 15-Oct-1999
C:Accession: S13512; S17073
R:Padanilam, B.J.; McLeod, L.B.; Suzuki, H.; Solursh, M.
Nucleic Acids Res. 19, 395, 1991
A:Title: Nucleotide sequence of an isoform of chicken retinoic acid binding protein-1
A:Reference number: S13512; MUID:91195065; PMID:1849630
A:Accession: S13512
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-455 <PAD>
A:Cross-references: EMBL:X56674; NID:g63785; PIDN:CAA39997.1; PID:g63786
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1991.
R:Nohno, T.; Muto, K.; Noji, S.; Saito, T.; Taniguchi, S.
Biochim. Biophys. Acta 1089, 273-275, 1991
A:Title: Isoforms of retinoic acid receptor beta expressed in the chicken embryo.
A:Reference number: S16243; MUID:91274365; PMID:1647216
A:Accession: S17073
A:Molecule type: mRNA
A:Residues: 1-29, 'P', '31-80 <NOH>
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology
C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation; zli

	Qy	1	MYESVEVGGETPNPFLVDFYQNQRACLLPEKGLPAPGYSPTPLRTPUNWG--SNHSIET	58
		:	: : :	
		:	: : :	
	Db	1	MFDCMEALCMGPRQ--LYDVTNRG-ACMLURK-----ASPFYAGLGFPAWTGTSYRSVET	52
		:	: : :	
		:	: : :	
	Qy	59	QSSSSBEIVPSPPPPPPLRIIXKPCFVCODKSSGYHYGVSACEGCKGFFRRSIQKNMYT	118
		:	: : :	
		:	: : :	
	Db	53	QSTSSSEMYVSPSPPPPPPVRYKPCFVCODKSSGYHYGVSSCEGCKGFFRRSIQKNMYT	112
		:	: : :	
		:	: : :	

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2003, 06:37:58 ; Search time 24 Seconds
(without alignments)
560.261 million cell updates/sec

Title: US-09-691-220-2
Perfect score: 2437
Sequence: 1 MYESVEVGGPTNPFVLVDF.....SCSPSLSPSSNRSSPATHSP 457

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2127	87.3	462	2	US-08-592-383-2
2	2127	87.3	462	2	US-08-095-7288-4
3	2127	87.3	462	5	PCT-US92-02320A-4
4	2114	86.7	416	4	US-08-764-870-4
5	2114	86.7	416	4	US-08-980-115-4
6	2112	86.7	797	2	US-08-095-7288-2
7	2112	86.7	797	5	PCT-US92-02320A-2
8	2111	86.6	462	6	5171671-2
9	1816	74.5	403	2	US-08-592-383-4
10	1777	72.9	448	6	5223606-2
11	1662	68.2	454	4	US-08-764-870-5
12	1662	68.2	454	4	US-08-980-115-5
13	1662	68.2	454	6	5260432-2
14	1625	66.7	368	6	5223606-3
15	658	27.0	410	4	US-08-764-870-1
16	658	27.0	410	4	US-08-980-115-1
17	652	26.8	410	4	US-08-764-870-2
18	652	26.8	410	4	US-08-980-115-2
19	652	26.8	410	6	5438126-2
20	619.5	25.4	461	4	US-08-764-870-3
21	619.5	25.4	461	4	US-08-980-115-3
22	611	23.1	355	6	5223606-4
23	570	23.4	463	1	US-08-336-408B-6
24	570	23.4	463	5	PCT-US91-00399-6
25	562	23.1	460	1	US-08-342-411A-2
26	560	23.0	461	1	US-08-330-518-2
27	560	23.0	461	1	US-08-330-283-2

28	560	23.0	461	2	US-08-646-248-2	Sequence 2, Appli
29	560	23.0	461	5	PCT-US95-13924-2	Sequence 2, Appli
30	560	23.0	461	5	PCT-US95-13931-2	Sequence 2, Appli
31	554	22.7	447	1	US-08-373-935-1	Sequence 1, Appli
32	552.5	22.7	440	1	US-08-333-358-8	Sequence 8, Appli
33	552.5	22.7	440	1	US-08-463-694-8	Sequence 8, Appli
34	552.5	22.7	440	1	US-08-694-501-8	Sequence 8, Appli
35	549	22.5	446	1	US-08-776-844-2	Sequence 2, Appli
36	546	22.4	446	2	US-08-372-652-3	Sequence 3, Appli
37	546	22.4	446	5	PCT-US95-16311-3	Sequence 3, Appli
38	545.5	22.4	443	1	US-08-342-411A-4	Sequence 4, Appli
39	542	22.2	525	4	US-08-764-870-7	Sequence 7, Appli
40	542	22.2	525	4	US-08-980-115-7	Sequence 7, Appli
41	531	21.8	533	1	US-07-952-800-2	Sequence 2, Appli
42	526.5	21.6	446	1	US-07-952-800-4	Sequence 4, Appli
43	518.5	21.3	462	1	US-08-336-408B-2	Sequence 2, Appli
44	518.5	21.3	462	4	US-08-764-870-6	Sequence 6, Appli
45	518.5	21.3	462	4	US-08-980-115-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-592-383-2
; Sequence 2, Application US/08592383
; Patent No. 5830760
; GENERAL INFORMATION:
; APPLICANT: Tsai, S. and S.J. Collins
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Re
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage
; COMPUTER: IBM PC/386 Compatible
; OPERATING SYSTEM: MS-DOS 4.01
; SOFTWARE: Word for Windows 5.01-t
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,383
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/099,242
; FILING DATE: July 28, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Broderick, Thomas F.
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: FPCR-1-7190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; DESCRIPTION: sequence encoded by SEQ. ID. NO. 1
US-08-592-383-2

Query Match 87.3%; Score 2127; DB 2; Length 462;
Best Local Similarity 91.6%; Pred. No. 1.3e-189;
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;
Qy 7 VGGPTNPFVLVDFYQNRACLLPEKGLPAGPYSTPLRPLWNGSNHSIETQSSSSEI 66
Db 32 LGLSP-PGALTQLQHO-----LPVSG-YSTP-----SPATIEQSSSSEI 71

QY 67 VSPSPSPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKNMYTCHRDKNCI 126
Db 72 VSPSPSPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKNMYTCHRDKNCI 131
QY 127 INKVTNRCOYCLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPEVGELEIKV 186
Db 132 INKVTNRCOYCLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPEVGELEIKV 191
QY 187 RKAHQETFPALCOLGKYTTNNSEORVSLDLDLWDFSELSKCIITKTVFAKOLPGFTT 246
Db 192 RKAHQETFPALCOLGKYTTNNSEORVSLDLDLWDFSELSKCIITKTVFAKOLPGFTT 251
QY 247 LTIADQITLLKAAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGLTDLVFA 306
Db 252 LTIADQITLLKAAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGLTDLVFA 311
QY 307 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQDPDRVDMLOEPLLEALKYVVRKRPSRP 366
Db 312 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQDPDRVDMLOEPLLEALKYVVRKRPSRP 371
QY 367 HMPKMLKITDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDLSGQPGGGG 426
Db 372 HMPKMLKITDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDLSGQPGGGG 431
QY 427 RDGGGLPPPGSCPSLSPSSNRSSPATHSP 457
Db 432 RDGGGLAPPGGSCPSLSPSSNRSSPATHSP 462
RESULT 2
US-08-095-728B-4
; Sequence 4, Application US/08095728B
; Patent No. 5843642
; GENERAL INFORMATION:
; APPLICANT: DMITROVSKY, ETHAN
; APPLICANT: WARRELL JR, RAYMOND P
; APPLICANT: MILLER JR, WILSON H
; APPLICANT: FRANKEL, STANLEY
; TITLE OF INVENTION: METHODS FOR THE DETECTION AND
; TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; STREET: 1185 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,728B
; FILING DATE: 21-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/673,838
; FILING DATE: 22-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WHITE, JOHN P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 38694-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-08-095-728B-4
Query Match 87.3%; Score 2127; DB 2: Length 462;
Best Local Similarity 91.6%; Pred. No. 1.3e-189;
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;
QY 7 VGGTTPNPFVLVDFYNQNRACLLPEKGLPAPGPYSTPLRTPLNWGSNNHSIETQSSSSEI 66
Db 32 LGGILSP-PGALTTLQHQ-----LPVSG-YSTP-----SPATIEQTSSSSEI 71
QY 67 VPSPSPSPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKNMYTCHRDKNCI 126
Db 72 VPSPSPSPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKNMYTCHRDKNCI 131
QY 127 INKVTNRCOYCLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPEVGELEIKV 186
Db 132 INKVTNRCOYCLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPEVGELEIKV 191
QY 187 RKAHQETFPALCOLGKYTTNNSEORVSLDLDLWDFSELSKCIITKTVFAKOLPGFTT 246
Db 192 RKAHQETFPALCOLGKYTTNNSEORVSLDLDLWDFSELSKCIITKTVFAKOLPGFTT 251
QY 247 LTIADQITLLKAAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGLTDLVFA 306
Db 252 LTIADQITLLKAAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGLTDLVFA 311
QY 307 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQDPDRVDMLOEPLLEALKYVVRKRPSRP 366
Db 312 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQDPDRVDMLOEPLLEALKYVVRKRPSRP 371
QY 367 HMPKMLKITDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDLSGQPGGGG 426
Db 372 HMPKMLKITDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDLSGQPGGGG 431
QY 427 RDGGGLPPPGSCPSLSPSSNRSSPATHSP 457
Db 432 RDGGGLAPPGGSCPSLSPSSNRSSPATHSP 462
RESULT 3
PCT-US92-02320A-4
; Sequence 4, Application PC/TUS9202320A
; GENERAL INFORMATION:
; APPLICANT: Sloan-Kettering Institute, For Cancer Research
; TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM
; STREET: 30 ROCKEFELLER PLAZA
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/02320A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 675,084
; FILING DATE: 22-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 673,838
; FILING DATE: 22-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WHITE, JOHN P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 38694-PCT

```

RESULT 4
US-08-764-870-4
; Sequence 4, Application US/08764870
; Patent No. 6236946
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S
; APPLICANT: Baxter, John D
; APPLICANT: Fletterick, Robert J
; APPLICANT: Wagner, Richard L
; APPLICANT: Kushner, Peter J
; APPLICANT: Arpilletti, James W
; APPLICANT: West, Brian
; TITLE OF INVENTION: Nuclear Rec
; TITLE OF INVENTION: Binding Dom
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Cooley Godward
; STREET: Five Palo Alto Square
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

RESULT 5
US-08-980-115-4
: Sequence 4, Application US/08980115
: Patent No. 6256622
: GENERAL INFORMATION:
: APPLICANT: Scanlan, Thomas S.
: APPLICANT: Baxter, John D.
: APPLICANT: Fletcherick, Robert J.
: APPLICANT: Wagner, Richard L.

APPLICANT: Kushner, Peter J.
APPLICANT: Apriletti, James W.
APPLICANT: West, Brian L.
APPLICANT: Shlau, Andrew K.
TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
FILE REFERENCE: UCAL-246/0205
CURRENT APPLICATION NUMBER: US/08/980,115
CURRENT FILING DATE: 1997-11-26
EARLIER APPLICATION NUMBER: 08/764,870
EARLIER FILING DATE: 1996-12-13
EARLIER APPLICATION NUMBER: 60/008,606
EARLIER FILING DATE: 1995-12-14
EARLIER APPLICATION NUMBER: 60/008,543
EARLIER FILING DATE: 1995-12-13
EARLIER APPLICATION NUMBER: 60/008,540
EARLIER FILING DATE: 1995-12-13
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 416
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (131)..(373)
OTHER INFORMATION: minimal ligand binding domain
US-08-980-115-4

Query Match 86.7%; Score 2114; DB 4; Length 416;
Best Local Similarity 97.1%; Pred. No. 1.8e-188;
Matches 402; Conservative 2; Mismatches 2; Indels 8; Gaps 1;

QY 52 SNH-----STETQSSSEIEIVSPSPPLPRIYKPCFVCDKSSGHHYGVSAEGC 103
DB 3 SNHVASGAGEAAIETQSSSEIEIVSPSPPLPRIYKPCFVCDKSSGHHYGVSAEGC 62
QY 104 KGFRRSIQNMVYCHRDKNCIINKVTRNPOCYCRLOKCFEVMGSKESVRNDRNKKKE 163
DB 63 KGFRRSIQNMVYCHRDKNCIINKVTRNPOCYCRLOKCFEVMGSKESVRNDRNKKKE 122
QY 164 VPKPESESTLTPVEGELIEKVRKAHQETFPALCQLGKVTNNSSQORVSLDIDLWDF 223
DB 123 VPKPESESTLTPVEGELIEKVRKAHQETFPALCQLGKVTNNSSQORVSLDIDLWDF 182
QY 224 SELSTKCIITVFPAKOLPGFTLTADQITLLKAACLDILILRICTRYTPEQDTWTFSD 283
DB 183 SELSTKCIITVFPAKOLPGFTLTADQITLLKAACLDILILRICTRYTPEQDTWTFSD 242
QY 284 GLTLNRTOMHNAFGPLTDLVFANOLLPLEMDDAETGLLSAICLCGDRDLEQPDV 343
DB 243 GLTLNRTOMHNAFGPLTDLVFANOLLPLEMDDAETGLLSAICLCGDRDLEQPDV 302
QY 344 DMLQEPLEALKVYVRRKRSRPHMPKMLKITDLRSISAKGAERVITLTKMEIPGSMPP 403
DB 303 DMLQEPLEALKVYVRRKRSRPHMPKMLKITDLRSISAKGAERVITLTKMEIPGSMPP 362
QY 404 LIQEMLENSEGLTSLQCPGGGDDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
DB 363 LIQEMLENSEGLTSLQCPGGGDDGGGLPPPGSCSPSLSPSSNRSSPATHSP 416

RESULT 6
US-08-095-7288-2
Sequence 2, Application US/08095728B
Patent No. 5843642
GENERAL INFORMATION:
APPLICANT: DMITROVSKY, ETHAN
APPLICANT: WARRELL JR, RAYMOND P
APPLICANT: MILLER JR, WILSON H
APPLICANT: FRANKEL, STANLEY
TITLE OF INVENTION: METHODS FOR THE DETECTION AND
TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)
NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:
ADDRESSEE: COOPER & DUNHAM LLP
STREET: 1185 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,728B
FILING DATE: 21-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/673,838
FILING DATE: 22-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WHITE, JOHN P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 38694-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-095-7288-2

Query Match 86.7%; Score 2112; DB 2; Length 797;
Best Local Similarity 99.3%; Pred. No. 6.9e-188;
Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 55 STETQSSSEIEIVSPSPPLPRIYKPCFVCDKSSGHHYGVSAEGCKGFFRRSIQKN 114
DB 395 AIETQSSSEIEIVSPSPPLPRIYKPCFVCDKSSGHHYGVSAEGCKGFFRRSIQKN 454
QY 115 MVYCHRDKNCIINKVTRNPOCYCRLOKCFEVMGSKESVRNDRNKKKEVPKPESESYT 174
DB 455 MVYCHRDKNCIINKVTRNPOCYCRLOKCFEVMGSKESVRNDRNKKKEVPKPESESYT 514
QY 175 LTPEVGELIEKVRKAHQETFPALCQLGKVTNNSSQORVSLDIDLWDFSELSTKCIIT 234
DB 515 LTPEVGELIEKVRKAHQETFPALCQLGKVTNNSSQORVSLDIDLWDFSELSTKCIIT 574
QY 235 VEFKALPGFTLTADQITLLKAACLDILILRICTRYTPEQDTWTFSDGLTLNRTOMHN 294
DB 575 VEFKALPGFTLTADQITLLKAACLDILILRICTRYTPEQDTWTFSDGLTLNRTOMHN 634
QY 295 AGFPLTDLVFANOLLPLEMDDAETGLLSAICLCGDRDLEQPDVDMLEQPLEAL 354
DB 635 AGFPLTDLVFANOLLPLEMDDAETGLLSAICLCGDRDLEQPDVDMLEQPLEAL 694
QY 355 KVVYKRRRSPRPHMPKMLKITDLRSISAKGAERVITLTKMEIPGSMPPLIQEMLENSE 414
DB 695 KVVYKRRRSPRPHMPKMLKITDLRSISAKGAERVITLTKMEIPGSMPPLIQEMLENSE 754
QY 415 LDTLSGQPGGGGDDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
DB 755 LDTLSGQPGGGGDDGGGLPPPGSCSPSLSPSSNRSSPATHSP 797

RESULT 7
PCT-US92-02320A-2
Sequence 2, Application PC/TUS9202320A
GENERAL INFORMATION:
APPLICANT: Sloan-Kettering Institute, For Cancer Research
TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOPER & DUNHAM
STREET: 30 ROCKEFELLER PLAZA
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/02320A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 675,084
FILING DATE: 22-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 673,838
FILING DATE: 22-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WHITE, JOHN P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 38694-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 644-0525
TELEX: (212) 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 797 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-02320A-2

Query Match 86.7%; Score 2112; DB 5; Length 797;

Best Local Similarity 99.3%; Pred. No. 6.9e-188;

Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 55 SIETQSSSEIIVSPSPPLPRIYPCFCVQDKSSGYHYGVSAACECKGFFRRSIQKN 114
:|||||
Db 395 AIETQSSSEIIVSPSPPLPRIYPCFCVQDKSSGYHYGVSAACECKGFFRRSIQKN 454
:|||||
QY 115 MYTCHRDKNCIINKVTNRCQYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPECSSEY 174
:|||||
Db 455 MYTCHRDKNCIINKVTNRCQYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPECSSEY 514
:|||||
QY 175 LTPVEGLIEKVRKAHQTETPALCOLGKYTTNNSSQVRSLDIDLWDFSELSTKCIKT 234
:|||||
Db 515 LTPVEGLIEKVRKAHQTETPALCOLGKYTTNNSSQVRSLDIDLWDFSELSTKCIKT 574
:|||||
QY 235 VEFKQLPGFTTLTADQITLLKAAACDLILRICTRYTPQDTMTFSDGLTLNRTOMHN 294
:|||||
Db 575 VEFKQLPGFTTLTADQITLLKAAACDLILRICTRYTPQDTMTFSDGLTLNRTOMHN 634
:|||||
QY 295 AGFGPLDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDQLEQPDVDMLOEPLEAL 354
:|||||
Db 635 AGFGPLDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDQLEQPDVDMLOEPLEAL 694
:|||||
QY 355 KVVYKRRPSRPHMFKMLMKITDLRSISAKGAERVITLKMIEPGSMPLPQIOMLENSEG 414
:|||||
Db 695 KVVYKRRPSRPHMFKMLMKITDLRSISAKGAERVITLKMIEPGSMPLPQIOMLENSEG 754
:|||||
QY 415 LDTLSGQPGGGGDDGLPPPGSCPSLSNSSRSPATHSP 457
:|||||
Db 755 LDTLSGQPGGGGDDGLPPPGSCPSLSNSSRSPATHSP 797
:|||||

RESULT 8

5171671-2
; Patent No. 5171671
; APPLICANT: EVANS, RONALD M.; ONG, ESTELITA S.; SEGUI,
; PRUDIMAR S.; THOMPSON, CATHERINE C.; UEMSONO, KAZUHIKO
; GUGUERE, VINCENT
; TITLE OF INVENTION: RETINOIC ACID RECEPTOR COMPOSITION
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/546,256
; FILING DATE: 06-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 276,536
; FILING DATE: 30-NOV-1988
; APPLICATION NUMBER: 128,331
; FILING DATE: 02-DEC-1987
; SEQ ID NO: 2:
; LENGTH: 462
5171671-2

Query Match 86.6%; Score 2111; DB 6; Length 462;

Best Local Similarity 91.1%; Pred. No. 3.9e-188;

Matches 411; Conservative 5; Mismatches 15; Indels 20; Gaps 4;

QY 7 VGGTNPFLVVDYFNQNRACLLPEKGLPAGPYSPLTRPLWNGSNHSTETQSSSEI 66
:|||||
Db 32 LGGLSP-FGALTTLQHQ-----LVSG-YSTP-----SPATETQSSSEI 71
:|||||
QY 67 VSPSPPLPRIYPCFCVQDKSSGYHYGVSAACECKGFFRRSIQKNMYTCHRDKNCI 126
:|||||
Db 72 VSPSPPLPRIYPCFCVQDKSSGYHYGVSAACECKGFFRRSIQKNMYTCHRDKNCI 131
:|||||
QY 127 INKVTNRCQYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPECSSEYTLTPVEGLIEKV 186
:|||||
Db 132 INKVTNRCQYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPECSSEYTLTPVEGLIEKV 191
:|||||
QY 187 RKAHQTETPALCOLGKYTTNNSSQVRSLDIDLWDFSELSTKCIKTVEFAKQLPGFTT 246
:|||||
Db 192 RKAHQTETPALCOLGKYTTNNSSQVRSLDIDLWDFSELSTKCIKTVEFAKQLPGFTT 251
:|||||
QY 247 LTIADQITLLKAAACDLILRICTRYTPQDTMTFSDGLTLNRTOMHNAFGPLTDLVFA 306
:|||||
Db 252 LTIADQITLLKAAACDLILRICTRYTPQDTMTFSDGLTLNRTOMHNAFGPLTDLVFA 311
:|||||
QY 307 FANQLLPLEMDDAETGLLSAICLICGDRDQLEQPDVDMLOEPLEALKYVYKRRPSR 366
:|||||
Db 312 FANQLLPLEMDDAETGLLSAICLICGDRDQLEQPDVDMLOEPLEALKYVYKRRPSR 371
:|||||
QY 367 HMFPMKMLKITDLRSISAKGAERVITLKMIEPGSMPLPQIOMLENSEGDLTSGQPGGG 426
:|||||
Db 372 HMFPMKMLKITDLRSISAKGAERVITLKMIEPGSMPLPQIOMLENSEGDLTSGQPGGG 431
:|||||
QY 427 RDGGLPPPGSCPSLSNSSRSPATHSP 457
:|||||
Db 432 RDGGLPPPGSCPSLSNSSRSPATHSP 462
:|||||

RESULT 9

US-08-592-383-4

; Sequence 4, Application US/08592383

; Patent No. 5830760

; GENERAL INFORMATION:

; APPLICANT: Tsai, S. and S.J. Collins

; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Res

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness

; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage

```

;SEQ ID NO: 2:
;      LENGTH: 448

```

APPLICATION NUMBER: US 60/008
FILING DATE: 13-DEC-1995

LENGTH: 448

```

> CURRENT APPLICATION NUMBER: US/08/980,115
> CURRENT FILING DATE: 1997-11-26
> EARLIER APPLICATION NUMBER: 08/764,870
> EARLIER FILING DATE: 1996-12-13
> EARLIER APPLICATION NUMBER: 60/008,606
> EARLIER FILING DATE: 1995-12-14
> EARLIER APPLICATION NUMBER: 60/008,543
> EARLIER FILING DATE: 1995-12-13
> EARLIER APPLICATION NUMBER: 60/008,540
> EARLIER FILING DATE: 1995-12-13
> NUMBER OF SEQ ID NOS: 17
> SOFTWARE: Patent Ver. 2.0
> SEQ ID NO 5
> LENGTH: 454
> TYPE: PRT
> ORGANISM: Homo sapiens
> FEATURE:
> NAME/KEY: DOMAIN
> LOCATION: (179)..(421)
> OTHER INFORMATION: minimal ligand binding
> US-08-980-115-5

```

```

Query Match          68.2%; Score 1662; DB 4; Length 454;
Best Local Similarity 72.8%; Pred. No. 2.3e-146;
Matches 319; Conservative 39; Mismatches 48; Indels 32; Gaps 4;

QY 30 PEKGLPAG-PYSTP-----LATPLWNG-----SNHSETQSSSEETVP 68
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 PGSGYGAGGFPFPGGALRSGPPPEMLSPSFRGLUGDPLPKEMASLSVETQSTSEEMVP 75
      : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 69 SPSPPLPLRIYPCFCYQDKSSGYHYGSACECKGFFRRSIQKNMYYTCHRDKNCIIN 128
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 76 SSPSPPPPPRYKPCFCVNDKSSGYHYGVSSCECKGFFRRSIQKNMYYTCHRDKNCIIN 135
      : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 129 KVTNRCQYCRQKCFEYGVGMSKESVRNDRNKKKEVPKPCSESYTLTPEVGEELIEKYRK 188
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 KVTNRCQYCRQKCFEYGVGMSKEAVRNDRNKKKEVKEEGSPDSYELSPLEELITKYSK 195
      : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 189 AHQETFPALCOLGKYTTNNSSQORVSLIDLWDFSELSTKCIITKVEFAKOLPGFTFLT 248
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 AHQETFPSCOLGKYTTNSSADHRVQLDGLWDFSELATKCIIKYVEFAKRLPGFTGLS 255
      : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 249 IADQITLLKAACLDILIRICTRYTPEDQTMTFSDGLTLNRTQMHNAGFGPLTDLVFAPA 308
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 IADQITLLKAACLDILMLRICTRYTPEDQTMTFSDGLTLNRTQMHNAGFGPLTDLVFAPA 315
      : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 309 NQLLPLEMDDAETGLLSAICLICGDRODLGPDQDRVMDLQEPLLAALKYVYKRRPSRPHM 368
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 GQLLPLEMDDTETGLLSAICLICGDRLDLEPEKVDKQLQEPLLAALRYARRRPSQPYM 375
      : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 369 FPKMLKITDLRSTISAGAERVITLKEIPEGMPPLIQEMLENSEGLDTLTSGQGP----- 423
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 FPRMLKITDLRGSTTGAERATILKMEIPEGMPPLIREMLENPEMFEDDSSQGPQHPNA 435
      : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 424 -----GGGRDGGGLPPP 435
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 SSEDVEPGGQGGGLKSP 453
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13
5260432-2
; Patent No. 5260432
; APPLICANT: TAKAKU, FUMIMARO; ISHIKAWA, TAKASHI; IMAWARI, MICHIO;
; EVANS, RONALD M.; UMESONO, KAZUHIKO
; TITLE OF INVENTION: HUMAN GAMMA RETINOIC ACID RECEPTOR DNA
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/370,407
; FILING DATE: 22-JUN-1989
; SEQ ID NO:2:
; LENGTH: 454
5260432-2

Query Match 68.2%; Score 1662; DB 6; Length 454;
Best Local Similarity 72.8%; Pred. No. 2.3e-146;
Matches 319; Conservative 39; Mismatches 48; Indels 32; Gaps 4;

QY 30 PEKGLPAG-PYSTP-----LRTPLWNG-----SNHSETQSSSEIYP 68
DB 16 PGSGYPCAGFPFAPGALRGSPFPEMLSPFRGLGQDLPKEMASLVETQSTSEEMVP 75
QY 69 SPSPPPPLRIYKPCFCODKSSGYHYGVSAECGCKGFFRRSIQKNMYVYCHRDKNKIIN 128
DB 76 SPSPPPPPRVYKPCFCVNDKSSGYHYGVSSCEGCKGFFRRSIQKNMYVYCHRDKNKIIN 135
QY 129 KVTNRNCOYCRLOKQCFEYGVSKESVNRDNKKKKEVPKPCSESYTLTPEVGELEIKVRK 188
DB 136 KVTNRNCOYCRLOKQCFEYGVSKESVNRDNKKKKEVPKPCSESYTLTPEVGELEIKVRK 195
QY 189 AHQETFPALCOLGKYTTNNSEORVSLDIDLWDKFSLSKCIKIVTVEFAKQLPGFTTLT 248
DB 196 AHQETFPALCOLGKYTTNNSEORVSLDIDLWDKFSLSKCIKIVTVEFAKQLPGFTTLT 255
QY 249 IADQITLLKAACLDILIRICTRYTPEDPTMTFSDGLTLNRTQHNAGFGLTDLVFAFA 308
DB 256 IADQITLLKAACLDILIRICTRYTPEDPTMTFSDGLTLNRTQHNAGFGLTDLVFAFA 315
QY 309 NQLPLEMDDAETGLLSAICLICGDRDLQPDQDVRDMLQEPLEALKVYVYKRRPSRPHM 368
DB 316 GOLLPLEMDDTETGLLSAICLICGDRDLQPDQDVRDMLQEPLEALKVYVYKRRPSRPHM 375
QY 369 FPKMLKITDLRISAKGAERTVTLKMEIPGSMPLQIEMLENSEGLDITLSGQPG----- 423
DB 376 FPKMLKITDLRISAKGAERTVTLKMEIPGSMPLQIEMLENSEGLDITLSGQPG----- 435
QY 424 -----GGGRDGGGLPPP 435
DB 436 SSEDEVPGGGKGLKSP 453

RESULT 14
5223606-3
; Patent No. 5223606
; APPLICANT: BLAUDIN DE THE, HUGHES, MARCHIO, AGNES, TIOLLAIS,
; PIERRE, DEJEAN, ANNE
; TITLE OF INVENTION: STEROID/THYROID HORMONE RECEPTOR-RELATED
; PROTEIN INAPPROPRIATELY EXPRESSED IN HUMAN HEPATOCELLULAR CARCINOMA
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/134,130
; FILING DATE: 17-DEC-1987
; PRIOR APPLICATION DATA:
; SEQ ID NO:3:
; LENGTH: 368
5223606-3

Query Match 66.7%; Score 1625; DB 6; Length 368;
Best Local Similarity 81.6%; Pred. No. 4.6e-143;
Matches 306; Conservative 27; Mismatches 32; Indels 10; Gaps 1;

QY 83 CFVQDKSSGYHYGVSAECGCKGFFRRSIQKNMYVYCHRDKNKIINKVTRNCPQYCRLOK 142
DB 1 CFVQDKSSGYHYGVSAECGCKGFFRRSIQKNMYVYCHRDKNKIINKVTRNCPQYCRLOK 60
QY 143 CFVGNKESVNRDNKKKKEVPKPCSESYTLTPEVGELEIKVRKAHOETFPALCOLGK 202
DB 61 CFVGNKESVNRDNKKKKEVPKPCSESYTLTPEVGELEIKVRKAHOETFPALCOLGK 120
QY 203 YTTNNSSEORVSLDIDLWDKFSLSKCIKIVTVEFAKQLPGFTTLTADQITLLKAACLD 262
DB 121 YTTNNSSEORVSLDIDLWDKFSLSKCIKIVTVEFAKQLPGFTTLTADQITLLKAACLD 180
QY 263 ILILIRICTRYTPEDPTMTFSDGLTLNRTQHNAGFGLTDLVFAFANQLLPLEMDDAETG 322
DB 181 ILILIRICTRYTPEDPTMTFSDGLTLNRTQHNAGFGLTDLVFAFANQLLPLEMDDAETG 240

QY 323 LLSAICLICGDRDLQPDQDVRDMLQEPLEALKVYVYKRRPSRPHMFKMLKITDLRSI 382
DB 241 LLSAICLICGDRDLQEPKVKYKLOEPLEALKVYVYKRRPSRPHMFKMLKITDLRSI 300
QY 383 SAKGAERTVTLKMEIPGSMPLQIEMLENSEGLDITLSGQPGGGGRDGGGLPPPGSCSPS 442
DB 301 SAKGAERTVTLKMEIPGSMPLQIEMLENSEGLDITLSGQPGGGGRDGGGLPPPGSCSPS 350
QY 443 LSPSSNRSSPATHSP 457
DB 351 ISPSSVNSGVQSPP 365

RESULT 15
US-08-764-870-1
; Sequence 1, Application US/08764870
; Patent No. 6236946
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S
; APPLICANT: Baxter, John D
; APPLICANT: Fletcher, Robert J
; APPLICANT: Wagner, Richard L
; APPLICANT: Kushner, Peter J
; APPLICANT: Apriletti, James W.
; APPLICANT: West, Brian
; TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward
; STREET: Five Palo Alto Square, 3000 El Camino Real
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,870
FILING DATE: 13-DEC-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,540
FILING DATE: 13-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,543
FILING DATE: 13-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,605
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Nakamura, Jackie N
REGISTRATION NUMBER: 35,966
REFERENCE/DOCKET NUMBER: UCAL-246/01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)843-5000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-764-870-1

Query Match 27.0%; Score 658; DB 4; Length 410;
Best Local Similarity 38.1%; Pred. No. 5e-53;
Matches 139; Conservative 64; Mismatches 120; Indels 42; Gaps 7;

QY 83 CFVQDKSSGYHYGVSAECGCKGFFRRSIQKNM--VYCHRDKNKIINKVTRNCPQYCR 140

Db	53	CVVCGDKATGYHCITCEGCKGFFRTTQKNLHPTSYKYDSCCVIDKTRNCQOLGRF	112
QY	141	OKCFEYVGMKESVRND-----RNKKKEV-----PKPECSYTLTREV	179
Db	113	KCIAVGMAMDLVLDDSKRVAKRKLEQRNRERKEEMIRSLQQRPEP-----TPEE	164
QY	180	GELIEKVRKAHQETPPA---LCQLGKYTTNNSQRV-----SLDIDLMDKPSLSTK	229
Db	165	WDLIHVATZAHSTRNAQGSWHQKRFPLDDTIGQSPIVMPDGDQKVDLEAFSEFTKIITP	224
QY	230	CIITVEFAKQLPGFTTLTADQITLLAKACLDILILRICTRYTEQDTMTFSDGLTLNR	289
Db	225	AITRVVDFAKLPMFSELPCEDQIILLGCCHEIMSLRAAVRYDEPSTLILSGEMTVKR	284
QY	290	TMHNAAGFPLTDLVFAFANQLPLEMDAAETGLLSAICLTCGDRQDLQEQDPRVMDLQEP	349
Db	285	KOLKNGGLGVSDAIFELGKSLSAFNLLDDETVALLQAVLLASTDRSGLLCVDKTEKSEA	344
QY	350	LLEALKVYVYKRRRPSRPHMFPKMLAKITDLIRISAKGAERVITLKMETPGSM-PPLIQEM	408
Db	345	YLLAFEHYVYVYKRRRPHMFPKMLAKITDLIRISAKGAERVITLKMETPGSM-PPLIQEM	408
QY	409	LENSE 413	
Db	405	FEDQE 409	

Search completed: March 29, 2003, 09:44:44
Job time : 26 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	2437	100.0	457	23	AAO21489	Human nuclear horm
2	2127	87.3	462	13	AAO27534	Human alpha protei
3	2127	87.3	462	20	AAW81964	Human RAR-alpha pr
4	2124	87.2	462	16	AAW68023	RAR-alpha. Homo s
5	2119	87.0	462	10	AAAP90395	Retinoic acid rece
6	2114	86.7	416	20	AAV21636	Ligand binding dom
7	2114	86.7	547	22	ABG18858	Novel human diagno
8	2112	86.7	503	22	ABG20948	Novel human diagno
9	2112	86.7	668	22	ABG18862	Novel human diagno
10	2112	86.7	797	13	AAO27533	mv1/RAR-alpha fusi

ALIGNMENTS

New human nuclear hormone receptor proteins and nucleic acids, useful as models or targets for developing human therapeutic targets, and in

PT identifying therapeutic proteins and modulators of nuclear hormone
PT receptor expression
PS Claim 1; Fig 2; 73pp; English.
XX The invention relates to an isolated peptide of a novel human nuclear
CC hormone receptor with a fully defined sequence of 457 amino acids given
CC in the specification. The novel human nuclear hormone receptor peptides
CC and nucleic acids encoding them can be used as models for the development
CC of human therapeutic targets, aid in the identification of therapeutic
CC proteins, and serve as targets for the development of human therapeutic
CC agents that modulate nuclear hormone receptor activity in cells and
CC tissues that express the nuclear hormone receptor. The nucleic acids may
CC be used as a query sequence to perform searches against sequence
CC databases to identify family members or related sequences, as probes or
CC primers, to construct recombinant vectors, to identify compounds that
CC modulate nuclear hormone receptor nucleic acid expression, in gene
CC therapy, and as antisense constructs to control nuclear hormone receptor
CC gene expression in cells, tissues or organisms. The polypeptides can be
CC used to raise antibodies or to elicit an immune response, as a reagent in
CC assays designed to determine protein levels in biological fluids, as
CC markers for tissues in which a corresponding protein is expressed, to
CC identify a binding partner/ligand to develop a system for the
CC identification of inhibitors of the binding reaction, in drug screening
CC assays, and to identify compounds that modulate protein activity. This
CC sequence represents the human nuclear hormone receptor protein of the
XX invention.
XX
SQ Sequence 457 AA;

Query Match 100.0%; Score 2437; DB 23; Length 457;
Best Local Similarity 100.0%; Pred. No. 7.7e-201;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYEVEVGPTNPFLVDFYQNRACILPEKGLPAGPYSTPLRTPLNWSNHSSTQTS 60
Db 1 MYEVEVGPTNPFLVDFYQNRACILPEKGLPAGPYSTPLRTPLNWSNHSSTQTS 60
QY 61 SSSEIIVSPSPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRSIQKNMYTCH 120
Db 61 SSSEIIVSPSPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRSIQKNMYTCH 120
QY 121 RDKNCIINKVTRNPOCYCRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPEVG 180
Db 121 RDKNCIINKVTRNPOCYCRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPEVG 180
QY 181 ELIEKVRKAHQETFPALCOLGKVTYNNSSQORVSLDLDWDFSELSTKCIKTVEFAKQ 240
Db 181 ELIEKVRKAHQETFPALCOLGKVTYNNSSQORVSLDLDWDFSELSTKCIKTVEFAKQ 240
QY 241 LPGETTLTADQITLLKAACTDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGL 300
Db 241 LPGETTLTADQITLLKAACTDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGL 300
QY 301 TDLVAFANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVVRK 360
Db 301 TDLVAFANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVVRK 360
QY 361 RPSRPHFPKMLKITDLSRSISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSG 420
Db 361 RPSRPHFPKMLKITDLSRSISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSG 420
QY 421 QPGGGRGDGGGLPPPGGSCPSLSPSSNRSSPATHSP 457
Db 421 QPGGGRGDGGGLPPPGGSCPSLSPSSNRSSPATHSP 457

RESULT 2
ID AAR27534
XX AAR27534 standard; Protein; 462 AA.
AC AAR27534;
XX

DT 09-MAR-1993 (first entry)
XX
DE RAR-alpha protein.
XX
KW Retinoic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;
KW APL; translocation; chromosome 17; PCR; primer;
KW [t(15;17)(q21;q11-22)]; breakpoint; polymerase chain reaction.
XX
OS Synthetic.
XX
PN W09216660-A.
XX
PD 01-OCT-1992.
XX
PE 23-MAR-1992; 92WO-US02320.
XX
PR 22-MAR-1991; 91US-0673838.
PR 22-MAR-1991; 91US-0675084.
XX
(SLOK) SLOAN KETTERING INST CANCER.
XX
PI Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;
PI Warrell RP;
XX
DR WPI; 1992-349240/42.
DR N-PSDB; AAQ29338.
XX
PT Marker for acute promyelocytic leukaemia and other neoplasias -
PT comprising nucleic acid and encoded abnormal retinoic acid
PT receptor-alpha receptor
XX
PS Disclosure; Page 43-46; 84pp; English.
XX
CC The sequence given represents the amino acid sequence of the retinoic
CC acid receptor (RAR)-alpha protein. The gene encoding this protein is
CC disrupted in a translocation of a portion of the long arm of chromosome
CC 17 onto the long arm of chromosome 15 [t(15;17)(q21;q11-22)]. This
CC causes a fusion RAR-alpha and myl which is characteristic of acute
CC promyelocytic leukemia (APL). The breakpoint region has been cloned
CC and it has been shown that DNA rearrangements are clustered in the
CC region of the first intron of RAR-alpha. This sequence was isolated
CC by polymerase chain reaction (PCR). The primers used for amplification
CC of this sequence can also be used to amplify the translocated region.
XX
SQ Sequence 462 AA;
Query Match 87.3%; Score 2127; DB 13; Length 462;
Best Local Similarity 91.6%; Pred. No. 3.5e-174;
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;
QY 7 VGGPTNPFLVDFYQNRACILPEKGLPAGPYSTPLRTPLNWSNHSSTQSSSEI 66
Db 32 LGLSLP-PGALTTLQHQ-----LPVSG-YSTP-----SPATETQSSSEI 71
QY 67 VPSPSPPPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRSIQKNMYTCHRDNCI 126
Db 72 VPSPSPPPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRSIQKNMYTCHRDNCI 131
QY 127 INKVTNPOCYCRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPEVCELIEK 186
Db 132 INKVTNPOCYCRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPEVCELIEK 191
QY 187 RKAHQETFPALCOLGKVTYNNSSQORVSLDLDWDFSELSTKCIKTVEFAKOLPGFTT 246
Db 192 RKAHQETFPALCOLGKVTYNNSSQORVSLDLDWDFSELSTKCIKTVEFAKOLPGFTT 251
QY 247 LTIADQITLLKAACTDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGLTDLVFA 306
Db 252 LTIADQITLLKAACTDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGLTDLVFA 311
QY 307 FANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVVRKRSRP 366
Db 312 FANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVVRKRSRP 371

QY 367 HMFPMKMKITDLRSISAKGAERVITLKMIEIPGSMPLIQEMLENSEGLDITLSGQPGGGG 426
Db 372 HMFPMKMKITDLRSISAKGAERVITLKMIEIPGSMPLIQEMLENSEGLDITLSGQPGGGG 431
QY 427 RDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
Db 432 RDGGGLAPPPGSCSPSLSPSSNRSSPATHSP 462
RESULT 3
AAW81964
ID AAW81964 standard; Protein; 462 AA.
XX AC AAW81964;
XX DT 05-FEB-1999 (first entry)
XX DE Human RAR-alpha protein.
XX KW Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;
KW acute promyelocytic leukaemia; APL; t(15;17); translocation;
KW treatment; all-trans retinoic acid.
XX OS Homo sapiens.
XX PN US5843642-A.
XX PD 01-DEC-1998.
XX PF 21-JUL-1993; 93US-0095728.
XX PR 21-JUL-1993; 93US-0095728.
XX PR 22-MAR-1991; 91US-0673838.
XX PR 22-MAR-1991; 91US-0675084.
XX PA (SLOK) SLOAN KETTERING INST CANCER RES.
XX PI Dmitrovsky E, Frankel S, Miller WH, Warrell RP;
XX WPI; 1999-044563/04.
XX DR N-PSDB; AAV64991.
XX PT Diagnosis of acute promyelocytic leukaemia - by detecting nucleic
XX acid encoding abnormal retinoic acid receptor-alpha
XX PS Disclosure; Column 35-38; 38pp; English.
XX CC This sequence represents the human retinoic acid receptor alpha,
CC RAR-alpha which is used in a method for identifying a subject with acute
CC promyelocytic leukaemia (APL) resulting from a t(15;17) translocation who
CC will respond to treatment with all-trans retinoic acid. The protein can
CC also be used to identify a subject with indications of APL who will not
CC respond to treatment with all-trans retinoic acid.
XX SQ Sequence 462 AA;
Query Match 87.3%; Score 2127; DB 20; Length 462;
Best Local Similarity 91.6%; Pred. No. 3.5e-174;
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;
QY 7 VGGTPTNPFVLVDFYQNRACLLPEKGLPAPGPYSTPLRTPLWNGSNHSIETQSSSEI 66
Db 32 LGGLSP-PGALTTLTQH-----LPVSG-YSTP-----SPATIEQTSSSEI 71
QY 67 VPSPSPPLPRIYKPCFVQDKSSGTHYGVSAACEGCKGFFRRSIQKNMVTCHRDKNCI 126
Db 72 VPSPSPPLPRIYKPCFVQDKSSGTHYGVSAACEGCKGFFRRSIQKNMVTCHRDKNCI 131
QY 127 INKVTNRNCPQYCRLOKCFEYGMKSRESYVRNDRNKKKKEVPKPECSSEYTLTPEVGELIEKV 186
Db 132 INKVTNRNCPQYCRLOKCFEYGMKSRESYVRNDRNKKKKEVPKPECSSEYTLTPEVGELIEKV 191

QY 187 RKAHQTFPFCALCOLGKYTTNNSSEORVSLDLDLWDFSELSKCIKKTVEPAKOLPGFTT 246
Db 192 RKAHQTFPFCALCOLGKYTTNNSSEORVSLDLDLWDFSELSKCIKKTVEPAKOLPGFTT 251
QY 247 LTADQITLLKAACLDILILIRICTRYTPEQDTMTFSDDLTLNRTQMHNAGFGPLTDLVFA 306
Db 252 LTADQITLLKAACLDILILIRICTRYTPEQDTMTFSDDLTLNRTQMHNAGFGPLTDLVFA 311
QY 307 FANQLLPLEMDDAETGLLSAICLICGDRDLEQPDVDMLOEPLLEALKVYVRRKRSRP 366
Db 312 FANQLLPLEMDDAETGLLSAICLICGDRDLEQPDVDMLOEPLLEALKVYVRRKRSRP 371
QY 367 HMFPMKMKITDLRSISAKGAERVITLKMIEIPGSMPLIQEMLENSEGLDITLSGQPGGGG 426
Db 372 HMFPMKMKITDLRSISAKGAERVITLKMIEIPGSMPLIQEMLENSEGLDITLSGQPGGGG 431
QY 427 RDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
Db 432 RDGGGLAPPPGSCSPSLSPSSNRSSPATHSP 462
RESULT 4
AAR68023
ID AAR68023 standard; Protein; 462 AA.
XX AC AAR68023;
XX DT 01-SEP-1995 (first entry)
XX DE RAR-alpha.
XX KW RAR-alpha; retinoic acid receptor alpha; hematopoietic; stem cell;
KW differentiation.
XX OS Homo sapiens.
XX PN WO9504143-A.
XX PD 09-FEB-1995.
XX PF 28-JUL-1994; 94WO-US08450.
XX PR 28-JUL-1993; 93US-0099242.
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PI Collins SJ, Tsai S;
XX DR WPI; 1995-082227/11.
XX DR N-PSDB; AAQ81476.
XX PT New haematopoietic stem cell lines with specific differentiation
XX properties - made by transfected stem cells with nucleic acid
XX encoding dominant negative suppressor of the retinoic acid
XX receptor alpha, useful e.g. for haematopoietic reconstitution
XX PS Disclosure; Page 56-58; 100pp; English.
XX CC An RAR-alpha cDNA (given in AAQ81477) was prepared that contained a
XX truncation of sequences coding for the C-terminal 59 amino acids
XX and part the 3' UTR of the wild-type human sequence (AAQ81476).
XX This truncated cDNA, designated RAR-alpha-403, encodes a protein
XX (AAR68024) containing the N-terminus, DNA-binding domain and part of
XX the hormone-binding domain of RAR-alpha (AAR68023).
XX SQ Sequence 462 AA;
Query Match 87.2%; Score 2124; DB 16; Length 462;
Best Local Similarity 91.4%; Pred. No. 6.4e-174;
Matches 412; Conservative 6; Mismatches 13; Indels 20; Gaps 4;
QY 7 VGGTPTNPFVLVDFYQNRACLLPEKGLPAPGPYSTPLRTPLWNGSNHSIETQSSSEI 66
Db 32 LGGLSP-PGALTTLTQH-----LPVSG-YSTP-----SPATIEQTSSSEI 71

Db 32 LGLLSP-PGALTTLQHQ-----LPVSG-YSTP-----SPATITQSSSEI 71

QY 67 VSPSPPLPRLPYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKNMVTCHRDKNCI 126

Db 72 VSPSPPLPRLPYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKNMVTCHRDKNCI 131

QY 127 INKVTNRCQYCLQKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPEVGELIEKV 186

Db 132 INKVTNRCQYCLQKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPEVGELIEKV 191

QY 187 RKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDFSELSKCIITVEFAKOLPGFTT 246

Db 192 RKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDFSELSKCIITVEFAKOLPGFTT 251

QY 247 LTIADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTNRTQMHNAGFGLTDLVFA 306

Db 252 LTIADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTNRTQMHNAGFGLTDLVFA 311

QY 307 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQPDVDMLOEPLEALKVYVYKRRPSRP 366

Db 312 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQPDVDMLOEPLEALKVYVYKRRPSRP 371

QY 367 HMFPMKMKITDLRSISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSGQPGGG 426

Db 372 HMFPMKMKITDLRSISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSGQPGGG 431

QY 427 RDGGGLPPPGSCSPSLPSSNRSSPATHSP 457

Db 432 RDGGGLAPPGGSCSPSLPSSNRSSPATHSP 462

RESULT 5

AAP90395

ID AAP90395 standard; protein; 462 AA.

XX AC AAP90395;

XX DT 01-NOV-1989 (first entry)

XX DE Retinoic acid receptor.

XX KW Clone phrAR1; retinoic acid receptor;

XX KW ligand complexes; human.

XX OS Homo sapiens (Human).

XX PN W08905355-A.

XX PD 15-JUN-1989.

XX PF 01-DEC-1988; 88WO-US04284.

XX PR 02-DEC-1987; 87US-0276536.

XX PA (SALK) SALK INST FOR BIOLOGICAL STUD.

XX PI Evans RM, Giguere V, Ong ES, Segui PS, Umesono K, Thompson CC;

XX DR WPI; 1989-192701/26.

XX DR N-PSDB; AAN90124.

XX PT DNA encoding retinoic acid receptor proteins

XX PT - used to produce proteins for studying complexes with

XX PT ligands and in diagnostic assays.

XX PS Disclosure; fig 1b; 74pp; English.

XX CC Primary sequence of a protein that has ligand binding and

XX CC transcription activating properties of retinoic acid receptor

XX CC (RAR) protein (see corresp. AAN90124). Used to make chimeric

XX CC receptors, to produce receptor, to study binding complexes,

XX CC and to screen cpds. for RAR-agonists and antagonists.

SQ Sequence 462 AA;

Query Match 87.0%; Score 2119; DB 10; Length 462;

Best Local Similarity 91.4%; Pred. No. 1.7e-173;

Matches 412; Conservative 5; Mismatches 14; Indels 20; Gaps 4;

QY 7 VGGPTNPFLVDFYQNRACLLPEKGLPAGPYSTLRLTPLWNGSNHSIETQSSSSEI 66

Db 32 LGLLSP-PGALTTLQHQ-----LPVSG-YSTP-----SPATITQSSSEI 71

QY 67 VSPSPPLPRLPYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKNMVTCHRDKNCI 126

Db 72 VSPSPPLPRLPYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKNMVTCHRDKNCI 131

QY 127 INKVTNRCQYCLQKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPEVGELIEKV 186

Db 132 INKVTNRCQYCLQKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPEVGELIEKV 191

QY 187 RKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDFSELSKCIITVEFAKOLPGFTT 246

Db 192 RKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDFSELSKCIITVEFAKOLPGFTT 251

QY 247 LTIADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTNRTQMHNAGFGLTDLVFA 306

Db 252 LTIADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTNRTQMHNAGFGLTDLVFA 311

QY 307 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQPDVDMLOEPLEALKVYVYKRRPSRP 366

Db 312 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQPDVDMLOEPLEALKVYVYKRRPSRP 371

QY 367 HMFPMKMKITDLRSISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSGQPGGG 426

Db 372 HMFPMKMKITDLRSISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSGQPGGG 431

QY 427 RDGGGLPPPGSCSPSLPSSNRSSPATHSP 457

Db 432 RDGGGLAPPGGSCSPSLPSSNRSSPATHSP 462

RESULT 6

AAY21636

ID AAY21636 standard; protein; 416 AA.

XX AC AAY21636;

XX DT 11-AUG-1999 (first entry)

XX DE Ligand binding domain of nuclear receptor hRARalpha.

XX KW Thyroid hormone receptor; aromatic compound; ligand binding domain;

XX KW alpha-glycerophosphate dehydrogenase; cardiac; obesity; triglyceride;

XX KW plasma cholesterol; anti-hypertriglyceridaemic; atherosclerosis; GPDH;

XX KW thyroid hormone replacement therapy; nuclear receptor.

XX OS Homo sapiens.

XX PN W0926966-A2.

XX PD 03-JUN-1999.

XX PF 25-NOV-1998; 98WO-US25296.

XX PR 26-NOV-1997; 97US-0980115.

XX PA (RECC) UNIV CALIFORNIA.

XX PI Aprilletti JW, Baxter JD, Fletterick RJ, Kushner PJ;

XX PI Scanlan TS, Shiau AK, Wagner RL, West BL;

XX DR WPI; 1999-357810/30.

XX PT Modulating activity of a thyroid hormone receptor


```

Db      542 PATHSP 547
|||||
RESULT 8
ABG20948
ID      ABG20948 standard; Protein; 563 AA.
XX      .
AC      ABG20948;
XX      .
DT      18-FEB-2002 (first entry)
XX      .
DE      Novel human diagnostic protein #20939.
XX      .
KW      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW      food supplement; medical imaging; diagnostic; genetic disorder.
XX      .
OS      Homo sapiens.
XX      .
PN      WO200175067-A2.
XX      .
PD      11-OCT-2001.
XX      .
PF      30-MAR-2001; 2001WO-US08631.
XX      .
PR      31-MAR-2000; 2000US-0540217.
PR      23-AUG-2000; 2000US-0649167.
XX      .
PA      (HYSE-) HYSEQ INC.
XX      .
PI      Drmanac RT, Liu C, Tang YT;
XX      .
DR      WPI; 2001-639362/73.
XX      .
DR      N-PSDB; AAS85135.
XX      .
PT      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity.
XX      .
PS      Claim 20; SEQ ID No 51307; 103pp; English.
XX      .
CC      The invention relates to isolated polynucleotide (I) and
CC      polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC      and gene mapping, and in recombinant production of (II). The
CC      polynucleotides are also used in diagnostics as expressed sequence tags
CC      for identifying expressed genes. (I) is useful in gene therapy techniques
CC      to restore normal activity of (II) or to treat disease states involving
CC      (II). (II) is useful for generating antibodies against it, detecting or
CC      quantitating a polypeptide in tissue, as molecular weight markers and as
CC      a food supplement. (II) and its binding partners are useful in medical
CC      imaging of sites expressing (II). (I) and (II) are useful for treating
CC      disorders involving aberrant protein expression or biological activity.
CC      The polypeptide and polynucleotide sequences have applications in
CC      diagnostics, forensics, gene mapping, identification of mutations
CC      responsible for genetic disorders or other traits to assess biodiversity
CC      and to produce other types of data and products dependent on DNA and
CC      amino acid sequences. ABG00010-ABG30377 represent novel human
CC      diagnostic amino acid sequences of the invention.
CC      Note: The sequence data for this patent did not appear in the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX      .
SQ      Sequence 563 AA;
Query Match      86.7%; Score 2112; DB 22; Length 563;
Best Local Similarity 99.3%; Pred. No. 9e-173;
Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy      55 SIETQSSSEIEVPPPPPLPRYKPCFVCQDKSSGHHYGVSAECGKGFFRSIQKN 114
:|||||
Db      161 AIETQSSSEIEVPPPPPLPRYKPCFVCQDKSSGHHYGVSAECGKGFFRSIQKN 220
|||||
115  MVTCHRDKNCIINKVTRNCOYCRLOKCFEYGMKSVNRNDRNKKKEVPKPECESEYT 174
|||||
221  MVTCHRDKNCIINKVTRNCOYCRLOKCFEYGMKSVNRNDRNKKKEVPKPECESEYT 280
|||||
175  LTPEVGELIEKVRKAHOETFPALCQLGKYTTNNSSQSRVSLDIDLWDFSELSTKCIIT 234
|||||
281  LTPEVGELIEKVRKAHOETFPALCQLGKYTTNNSSQSRVSLDIDLWDFSELSTKCIIT 340
|||||
235  VEFKQLPGFTTLLTADQITLLKAACTDILILRITCTRYTPEDQTMFTFSDGLTLNRTOMHN 294
|||||
341  VEFKQLPGFTTLLTADQITLLKAACTDILILRITCTRYTPEDQTMFTFSDGLTLNRTOMHN 400
|||||
295  AGFGPLTDLVFAFANQLLPLEMDDAETGLLSAICLCIGDRODLEOPDRVMDLQEPLEAL 354
|||||
401  AGFGPLTDLVFAFANQLLPLEMDDAETGLLSAICLCIGDRODLEOPDRVMDLQEPLEAL 460
|||||
355  KVVYRRRPSRPHMFPKMLKMTDLRSISAKGAERVITLKEIPGSMPLLIQEMLENSEG 414
|||||
461  KVVYRRRPSRPHMFPKMLKMTDLRSISAKGAERVITLKEIPGSMPLLIQEMLENSEG 520
|||||
415  LDTLSGQPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
|||||
521  LDTLSGQPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 563
|||||
RESULT 9
ABG18862
ID      ABG18862 standard; Protein; 668 AA.
XX      .
AC      ABG18862;
XX      .
DT      18-FEB-2002 (first entry)
XX      .
DE      Novel human diagnostic protein #18853.
XX      .
KW      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW      food supplement; medical imaging; diagnostic; genetic disorder.
XX      .
OS      Homo sapiens.
XX      .
PN      WO200175067-A2.
XX      .
PD      11-OCT-2001.
XX      .
PF      30-MAR-2001; 2001WO-US08631.
XX      .
PR      31-MAR-2000; 2000US-0540217.
PR      23-AUG-2000; 2000US-0649167.
XX      .
PA      (HYSE-) HYSEQ INC.
XX      .
PI      Drmanac RT, Liu C, Tang YT;
XX      .
DR      WPI; 2001-639362/73.
XX      .
DR      N-PSDB; AAS83049.
XX      .
PT      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity.
XX      .
PS      Claim 20; SEQ ID No 49221; 103pp; English.
XX      .
CC      The invention relates to isolated polynucleotide (I) and
CC      polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC      and gene mapping, and in recombinant production of (II). The
CC      polynucleotides are also used in diagnostics as expressed sequence tags
CC      for identifying expressed genes. (I) is useful in gene therapy techniques
CC      to restore normal activity of (II) or to treat disease states involving
CC      (II). (II) is useful for generating antibodies against it, detecting or
CC      quantitating a polypeptide in tissue, as molecular weight markers and as

```

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 668 AA;

Query Match 86.7%; Score 2112; DB 22; Length 668;
 Best Local Similarity 99.3%; Pred. No. 1.1e-172;
 Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 55 SIETQSSSEIIVPSPPPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKN 114
 Db :|||||
 266 AIETQSSSEIIVPSPPPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKN 325
 QY 115 MYVTCHRDKNCIINKVTNRNCOYCRLOKCFEVMGSKESVRNDRNKKKEVPKPECSSEY 174
 Db :|||||
 326 MYVTCHRDKNCIINKVTNRNCOYCRLOKCFEVMGSKESVRNDRNKKKEVPKPECSSEY 385
 QY 175 LPEVGELEIEKVRKAHQETFPALCOLGKYTTNNSSQRYSLDIDLWDFSELSTKCIIT 234
 Db :|||||
 386 LPEVGELEIEKVRKAHQETFPALCOLGKYTTNNSSQRYSLDIDLWDFSELSTKCIIT 445
 QY 235 VEFKQLPGFTTLTIADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTOMHN 294
 Db :|||||
 446 VEFKQLPGFTTLTIADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTOMHN 505
 QY 295 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRQDLEQDPRVDMQLPELLEAL 354
 Db :|||||
 506 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRQDLEQDPRVDMQLPELLEAL 565
 QY 355 KYVVRKRPSRPHMFKMLKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEG 414
 Db :|||||
 566 KYVVRKRPSRPHMFKMLKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEG 625
 QY 415 LDTLSQPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
 Db :|||||
 626 LDTLSQPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 668

RESULT 10
 AAR27533
 ID AAR27533 standard; Protein; 797 AA.

XX AC AAR27533;

DT 09-MAR-1993 (first entry)

DE DE myl/RAR-alpha fusion.

KW Retinoic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;
 KW APL; translocation; chromosome 17; chromosome 15; PCR; primer;
 KW [t(15;17)(q21;q11-22)]; breakpoint; polymerase chain reaction.

OS Synthetic.

PN W09216660-A.

PD 01-OCT-1992.

PF 23-MAR-1992; 92WO-US02320.

XX 22-MAR-1991; 91US-0673838.

PR 22-MAR-1991; 91US-0675084.

XX

PA (SLOK) SLOAN KETTERING INST CANCER.
 XX Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;
 PI Warrell RP;
 XX WPI; 1992-349240/42.
 DR N-PSDB; AAQ29334.
 XX Marker for acute promyelocytic leukaemia and other neoplasias
 PT comprising nucleic acid and encoded abnormal retinoic acid
 PT receptor-alpha receptor
 XX Disclosure; Fig 4; 84pp; English.

XX SQ Sequence 797 AA;

Query Match 86.7%; Score 2112; DB 13; Length 797;
 Best Local Similarity 99.3%; Pred. No. 1.5e-172;
 Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 55 SIETQSSSEIIVPSPPPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKN 114
 Db :|||||
 395 AIETQSSSEIIVPSPPPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKN 454
 QY 115 MYVTCHRDKNCIINKVTNRNCOYCRLOKCFEVMGSKESVRNDRNKKKEVPKPECSSEY 174
 Db :|||||
 455 MYVTCHRDKNCIINKVTNRNCOYCRLOKCFEVMGSKESVRNDRNKKKEVPKPECSSEY 514
 QY 175 LPEVGELEIEKVRKAHQETFPALCOLGKYTTNNSSQRYSLDIDLWDFSELSTKCIIT 234
 Db :|||||
 515 LPEVGELEIEKVRKAHQETFPALCOLGKYTTNNSSQRYSLDIDLWDFSELSTKCIIT 574
 QY 235 VEFKQLPGFTTLTIADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTOMHN 294
 Db :|||||
 575 VEFKQLPGFTTLTIADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTOMHN 634
 QY 295 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRQDLEQDPRVDMQLPELLEAL 354
 Db :|||||
 635 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRQDLEQDPRVDMQLPELLEAL 694
 QY 355 KYVVRKRPSRPHMFKMLKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEG 414
 Db :|||||
 695 KYVVRKRPSRPHMFKMLKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEG 754
 QY 415 LDTLSQPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
 Db :|||||
 755 LDTLSQPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 797

RESULT 11

AAW81963

ID AAW81963 standard; Protein; 797 AA.

XX AC AAW81963;

DT 05-FEB-1999 (first entry)

DE DE Human myl/RAR-alpha fusion protein.

XX Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;

KW acute promyelocytic leukaemia; APL; t(15;17); translocation;

KW treatment; all-trans retinoic acid.

XX

OS Homo sapiens.
XX Synthetic.
PN US5843642-A.
XX
PD 01-DEC-1998.
XX
PF 21-JUL-1993; 93US-0095728.
XX
PR 21-JUL-1993; 93US-0095728.
PR 22-MAR-1991; 91US-0673838.
XX 22-MAR-1991; 91US-0675084.
XX
PA (SLOAN) SLOAN KETTERING INST CANCER RES.
XX
PI Dmirovsy E, Frankel S, Miller WH, Warrell RP;
XX WPI; 1999-044563/04.
DR N-PSDB; AAV64990.
XX
XX Diagnosis of acute promyelocytic leukaemia - by detecting nucleic
PT acid encoding abnormal retinoic acid receptor-alpha
XX
PS Disclosure; Fig 4A-G; 38pp; English.
XX
CC This sequence represents a human myl/RAR-alpha (retinoic acid receptor
CC alpha) fusion protein which is used in a method for identifying a
CC subject with acute promyelocytic leukaemia (APL) resulting from a
CC t(15;17) translocation who will respond to treatment with all-trans
CC retinoic acid. The protein can also be used to identify a subject
CC with indications of APL who will not respond to treatment with all-trans
CC retinoic acid.
XX
XX Sequence 797 AA;
XX
Query Match 86.7%; Score 2112; DB 20; Length 797;
Best Local Similarity 99.3%; Pred. No. 1.5e-172;
Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 55 SIFQSSSEIEIVSPSPPLRIYKPCVFCODKSSGYHYGVSAEGCKGFFRRSIQKN 114
Db 395 ALETQSSSEIEIVSPSPPLRIYKPCVFCODKSSGYHYGVSAEGCKGFFRRSIQKN 454
QY 115 MYTCHRDNCIINKVTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 174
Db 455 MYTCHRDNCIINKVTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 514
QY 175 LTPEVGELIEKVKRAHQETFPALCOLGKYTTNNSSQYRVSLLDIDLWDFSELSTKCIKT 234
Db 515 LTPEVGELIEKVKRAHQETFPALCOLGKYTTNNSSQYRVSLLDIDLWDFSELSTKCIKT 574
QY 235 VEFKQLPGFTLTADQITLLKAACLDLILIRICTRYTPEQDTWTFSDGLTLNRTOMHN 294
Db 575 VEFKQLPGFTLTADQITLLKAACLDLILIRICTRYTPEQDTWTFSDGLTLNRTOMHN 634
QY 295 AGFGLTDLVFAFANQLPLEMDDAETGLLSAICLCIGDRQDLEQPDVDMQEPLEAL 354
Db 635 AGFGLTDLVFAFANQLPLEMDDAETGLLSAICLCIGDRQDLEQPDVDMQEPLEAL 694
QY 355 KYVYKRRPSRPHMPKMLMTKTLRSISAKGAERVITLTKMEIPGSMPLIOEMLENSEG 414
Db 695 KYVYKRRPSRPHMPKMLMTKTLRSISAKGAERVITLTKMEIPGSMPLIOEMLENSEG 754
QY 415 LDTLSGQPGGGGDDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
Db 755 LDTLSGQPGGGGDDGGGLPPPGSCSPSLSPSSNRSSPATHSP 797
RESULT 12
ABG18860
ID ABG18860 standard; Protein; 819 AA.
XX
AC ABG18860;

XX 18-FEB-2002 (first entry)
DT Novel human diagnostic protein #18851.
XX
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
KW
KW Homo sapiens.
XX
OS WO200175067-A2.
XX
PN 11-OCT-2001.
XX
PD 30-MAR-2001; 2001WO-US08631.
XX
PF 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
PI N-PSDB; AAS83047.
XX
DR New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX Claim 20; SEQ ID NO 49219; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 819 AA;
XX
Query Match 86.7%; Score 2112; DB 22; Length 819;
Best Local Similarity 99.3%; Pred. No. 1.5e-172;
Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 55 SIFQSSSEIEIVSPSPPLRIYKPCVFCODKSSGYHYGVSAEGCKGFFRRSIQKN 114
Db 417 ALETQSSSEIEIVSPSPPLRIYKPCVFCODKSSGYHYGVSAEGCKGFFRRSIQKN 476
QY 115 MYTCHRDNCIINKVTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 174
Db 477 MYTCHRDNCIINKVTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 536
QY 175 LTPEVGELIEKVKRAHQETFPALCOLGKYTTNNSSQYRVSLLDIDLWDFSELSTKCIKT 234
Db 537 LTPEVGELIEKVKRAHQETFPALCOLGKYTTNNSSQYRVSLLDIDLWDFSELSTKCIKT 596

```

QY 235 VEFKQLPGFTTITADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTQMHN 294
Db 597 VEFKQLPGFTTITADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTQMHN 656
QY 295 AGFGPLTDLVFAFANQLLPLEMDAETGLLSAICLCGDRQDLQPDVDMQLQEPLEAL 354
Db 657 AGFGPLTDLVFAFANQLLPLEMDAETGLLSAICLCGDRQDLQPDVDMQLQEPLEAL 716
QY 355 KVVYKRRSPRHPMPKMLKTTDLRSISAKGAERVITLKMETPGSMPLIOEMLENSEG 414
Db 717 KVVYKRRSPRHPMPKMLKTTDLRSISAKGAERVITLKMETPGSMPLIOEMLENSEG 776
QY 415 LDTLSGQPGGGRDGLPPPGSCPSLSPSSNRSSPATHSP 457
Db 777 LDTLSGQPGGGRDGLPPPGSCPSLSPSSNRSSPATHSP 819

RESULT 13
AAR10547
ID AAR10547 standard; protein; 462 AA.
AC AAR10547;
XX
DT 11-APR-1991 (first entry)
DE Murine Retinoic Acid Receptor-alpha (deduced from cDNA).
KW mouse; retinoic acid receptor; mRAR-alpha; embryogenesis.
OS Mus musculus.
FH Key
FT Location/Qualifiers
FT 1..59
FT /label= Region A
FT /note= "important for differential trans-acti-
FT tion by other nuclear receptors"
FT Region
FT 60..87
FT /label= Region B
FT /note= "well conserved between mRAR's"
FT Region
FT 88..153
FT /label= Region C
FT /note= "DNA binding domain. 95 per cent amino acid
FT identity between mRAR's"
FT Region
FT 154..199
FT /label= Region D
FT /note= "putative hinge region"
FT Region
FT 200..419
FT /label= Region E
FT /note= "ligand binding domain"
FT Region
FT 420..462
FT /label= Region F
FT XX
FT EP411323-A.
FT PN
FT XX
FT 06-FEB-1991.
FT PD
FT XX
FT 29-JUN-1990; 90EP-0112469.
FT PF
FT XX
FT 29-MAR-1990; 90US-0502140.
FT PR
FT 30-JUN-1989; 89US-0374690.
FT XX
FT (INRM ) INSERM INST NAT SANTE.
FT PA (SQUI ) LES LABS SQUIBB SA.
FT PA
FT XX
FT Chambon P, Kastner P, Krust A, Petkovich M, Zelent A;
FT PI Leroy P, Mendelson C, Staub A;
FT XX
FT WP1; 1991-038271/06.
FT DR
FT XX
FT Novel human and mouse retinoic acid receptors - encode proteins
FT used to assay for agonists and antagonists
FT XX
FT Disclosure; Fig 1; 33pp; English.
FT PS

```

```

XX An 11.5 day-old total mouse embryo lambda gt10 cDNA library was
CC screened with labelled human RAR-alpha probes. Positive clones were
CC isolated, mapped and sequenced. A set of mRAR-alpha clones was
CC identified on the basis of a 98 percent homology of their cDNA-
CC deduced amino acid sequence with that of human RAR-alpha. There are
CC 8 amino acid substitutions (mostly conservative) between the human
CC and mouse sequences. See also AAR10548, AAR10388-9 and AAR10405-8.
XX
SQ Sequence 462 AA;

Query Match 85.7%; Score 2089; DB 12; Length 462;
Best Local Similarity 90.0%; Pred. No. 6 5e-171;
Matches 406; Conservative 8; Mismatches 17; Indels 20; Gaps 4;

QY 7 VGGFTNPFLVDFYNQNRACLLPEKGLPAPGPYSTPLRTPLMNGSNHSITQSSSBEI 66
Db 32 LGGLSP-PGALTSLQHQ-----LPVSG-YSTP-----SPATLETQSSSBEI 71

QY 67 VPSPPPPLPRIYKPCFVCQDKSSGYHYGSACEGCKGFRRSIQKNMYTCHRDKNCI 126
Db 72 VPSPPPPLPRIYKPCFVCQDKSSGYHYGSACEGCKGFRRSIQKNMYTCHRDKNCI 131

QY 127 INKVTNRPCQYCRLOKCFEYVGMSESVNRDNKKKKKVPKPECSYTLTPEVGELIEKV 186
Db 132 INKVTNRPCQYCRLOKCFEYVGMSESVNRDNKKKKKVPKPECSYTLTPEVGELIEKV 191

QY 187 RKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDKFESELSKCIITVEFAKOLPGFTT 246
Db 192 RKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDKFESELSKCIITVEFAKOLPGFTT 251

QY 247 LTIADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTQMHNAGPGLTDLVFA 306
Db 252 LTIADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTQMHNAGPGLTDLVFA 311

QY 307 FANQLLPLEMDAETGLLSAICLCGDRQDLQPDVDMQLQEPLEALKVYVYKRRSPSRP 366
Db 312 FANQLLPLEMDAETGLLSAICLCGDRQDLQPDVDMQLQEPLEALKVYVYKRRSPSRP 371

QY 367 HMFPMKMLKITDLRSISAKGAERVITLKMETPGSMPLIOEMLENSEGDLTSLGQPGGG 426
Db 372 HMFPMKMLKITDLRSISAKGAERVITLKMETPGSMPLIOEMLENSEGDLTSLGQSGGGT 431

QY 427 RDGGGLPPPGSCPSLSPSSNRSSPATHSP 457
Db 432 RDGGGLPPPGSCPSLSPSSNRSSPATHSP 462

RESULT 14
AAR84724
ID AAR84724 standard; Protein; 462 AA.
AC AAR84724;
XX
DT 13-JUN-1996 (first entry)
DE Murine retinoic acid receptor alpha.
KW Retinoic acid receptor; mouse; RAR; RAR-gamma; transcription factor;
KW skin; RAR-alpha.
XX Mus musculus.
XX OS
XX Key
XX Location/Qualifiers
FT Region
FT 1..59
FT /note= "A region"
FT Region
FT 60..87
FT /note= "B region"
FT Region
FT 88..153
FT /note= "C region"
FT Region
FT 154..199
FT /note= "D region"
FT Region
FT 200..419

```


Query Match 83.2%; Score 2026.5; DB 22; Length 470;
Best Local Similarity 88.5%; Pred. No. 1.6e-165;
Matches 402; Conservative 7; Mismatches 22; Indels 23; Gaps 7;

QY	7	VGGPTNPFLVVDYFNONRACLPEKGLPAGPYSTPLRPLWNGSNHSTETOSSSEI	66
Db	37	LGGLSP-PCALTTLQHQ-----LPVSG-YSTP-----SPATETOSSSEI	76
QY	67	VSPSPSPPLPRIYKPCFYCDKSSGYHYGSACEGCKGFFRRSIQKNMYTCHRDKNCI	126
Db	77	VSPSPSPPLPRIYKPCFYCDKSSGYHYGSACEGCKGFFRRSIQKNMYTCHRDKNCI	136
QY	127	INKVTRNPOYCRLOKCFEYVGSKEVSRNDRNKKKEVPKPECSSEYTLTPEVGELIEKV	186
Db	137	INKVTRNRCQYCRLOKCFEYVGSKEVSRNDRNKKKEVPKPECSSEYTLTPEVGELIEKV	196
QY	187	RKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDKFSELSTKCIITKTVFAK-QLPGFT	245
Db	197	RKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDKFSELSTKCIITKNGEGROGLPGFT	256
QY	246	TLTADQITLLKAACLDIILIRICTRYTPQDWTFS-DGLTLNRTOMHNAGFGLTDLV	304
Db	257	TLTADQITLLKAACLDIILIRICTRYTPAGTTPXPSWDGLTLNRTOMHNAGFGLTDLV	316
QY	305	FAFANQLLPLEMDA-ETGLLSAICLCGRDQDLEQPDVDMLEQLLEALKVYVRKRRP	363
Db	317	FAFANQLLPLEMDARETGLLSAICLCGRDQDLEQPDVDMLEQLLEALKVYVRKRRP	376
QY	364	SRPHMFPKMLKITDLRSISAKGAERVITLKMEIPGSMPLIOEMLENSEGLDTLSCQPG	423
Db	377	SRPHMFPKMLKITDLRSISAKGAERVITLKMEIPGSMPLIOEMLENSEGLDTLSCQPG	436
QY	424	GGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP	457
Db	437	GGGRDGGGLAPPGSCSPSLSPSSNRSSPATHSP	470

Search completed: March 29, 2003, 06:38:55
Job time : 58 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2127	87.3	462	10	US-09-797-727-3		Sequence 3, Appli
2	2124	87.2	462	9	US-09-814-604-1		Sequence 1, Appli
3	1771	72.7	448	9	US-09-814-604-2		Sequence 2, Appli
4	1765	72.4	448	10	US-09-797-727-4		Sequence 4, Appli
5	1662	68.2	454	9	US-09-814-604-3		Sequence 3, Appli
6	1662	68.2	454	10	US-09-797-727-2		Sequence 2, Appli
7	560	23.0	461	12	US-10-013-823-3		Sequence 3, Appli
8	549	22.5	446	10	US-09-909-446-2		Sequence 2, Appli
9	549	22.5	446	10	US-09-909-325-2		Sequence 2, Appli
10	549	22.5	446	10	US-09-909-326-2		Sequence 2, Appli
11	545	22.4	446	12	US-10-013-823-2		Sequence 2, Appli
12	519.5	21.3	598	10	US-09-853-386-32		Sequence 32, Appl
13	519.5	21.3	598	10	US-09-853-386-35		Sequence 35, Appl
14	519.5	21.3	598	10	US-09-853-386-36		Sequence 36, Appl
15	519.5	21.3	598	10	US-09-853-386-37		Sequence 37, Appl
16	519.5	21.3	598	10	US-09-853-386-38		Sequence 38, Appl
17	519.5	21.3	598	10	US-09-853-386-39		Sequence 39, Appl
18	519.5	21.3	598	10	US-09-853-386-40		Sequence 40, Appl
19	519.5	21.3	598	10	US-09-853-386-41		Sequence 41, Appl

4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
8


```
; TITLE OF INVENTION: GAMMA RETINOIC ACID RECEPTOR
; FILE REFERENCE: SALK1150-3
; CURRENT APPLICATION NUMBER: US/09/797,727
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 08/486,325
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/100,039
; PRIOR FILING DATE: 1993-07-30
; PRIOR APPLICATION NUMBER: PCT/US90/03564
; PRIOR FILING DATE: 1990-06-22
; PRIOR APPLICATION NUMBER: US 07/370,407
; PRIOR FILING DATE: 1989-06-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Human
; NAME/KEY: misc.feature
; OTHER INFORMATION: Human Retinoic Acid Receptor-beta (hRAR-beta)
US-09-797-727-4
```

```
Query Match 72.4%; Score 1765; DB 10; Length 448;
Best Local Similarity 74.1%; Pred. No. 2.8e-136;
Matches 341; Conservative 43; Mismatches 58; Indels 18; Gaps 5;

QY 1 MYSEVEGGPTNPFVVDYFNON--RACLLPEKGLPAPGPTSTPLRTPLWNGSN--HSIE 57
DB 1 MFDGMDVLSVSPGQ--ILDFTASPSQMLQEKALKA--CFSGLTQTEWQHRHTAQSTE 55

QY 58 TQSSSEIEVPSPPLPRIYKPCFCQDKSSGYHYGVSAECGCKGFFRRSIQKNMYY 117
DB 56 TQTSSEELVSPSPPLPPRYKPCFCQDKSSGYHYGVSAECGCKGFFRRSIQKNMYY 115

QY 118 TCHRDKNCKVINKVTRNCOYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTTP 177
DB 116 TCHRDKNCKVINKVTRNCOYCRLOKCFEVMGSKESVRNDRNKKKEVSKQECTSEYMTA 175

QY 178 EVGELIEKVRKAHQETPPALCQLGKYYTNNSSQORVSLDIDLMDKFSSELSTKCIITKTV 237
DB 176 ELDDLTEKIRKAHQETPPSLCQLGKYYTNNSSADHRVRLDLGLMDKFSSELATKCIIVEF 235

QY 238 AKQLPGFTTLTIADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTLNRTQHNAGF 297
DB 236 AKRLPGFTGLSIADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTLNRTQHNAGF 295

QY 298 GPLDLVFAPANQLLPLEMDDAETGLLSAICLCGDRDLQOPDRVDMLOEPLEALKVY 357
DB 296 GPLDLVFTFANQLLPLEMDDDTETGLLSAICLCGDRDLQOPDRVDMLOEPLEALKVY 355

QY 358 VRRRRSRPHFPKMLKITDLRSISAKGAERVITLKEIPGSMPLLIQEMLENSEGLDT 417
DB 356 IRKRSPKPHFPKMLKITDLRSISAKGAERVITLKEIPGSMPLLIQEMLENSEGLDT 415

QY 418 LSGQGGGGRDGGGLPPPPGSCSLSLSPSSNRKSPATHSP 457
DB 416 LTPSSSGNTAEH-----SPSISPSSVNSGVQSP 445
```

```
RESULT 5
US-09-814-604-3
; Sequence 3, Application US/09814604
; Publication No. US20030003517A1
; GENERAL INFORMATION:
; APPLICANT: Klein, Elliott S.
; APPLICANT: Chandraratna Roshantha A.
; TITLE OF INVENTION: Methods of Detecting Dissociated Nuclear
; TITLE OF INVENTION: Hormone Receptor Ligands
; FILE REFERENCE: P-AR 4528
; CURRENT APPLICATION NUMBER: US/09/814,604
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 52
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-814-604-3

Query Match 68.2%; Score 1662; DB 9; Length 454;
Best Local Similarity 72.8%; Pred. No. 7.5e-128;
Matches 319; Conservative 39; Mismatches 48; Indels 32; Gaps 4;

QY 30 PEKGLPAPG-PYSTP-----LRTPLWNG-----SNHSIETQSSSEIIVP 68
DB 16 PGSGYPGAGFPFAFGALRGSPFFEMLSFSGGLGQDPLPKEMASLSVETQSTSSSEMV 75

QY 69 SPPSPPLPRIYKPCFCVQDKSSGYHYGVSAECGCKGFFRRSIQKNMYYTCHRDKNCKI 128
DB 76 SSPSPPLPRIYKPCFCVQDKSSGYHYGVSAECGCKGFFRRSIQKNMYYTCHRDKNCKI 135

QY 129 KYTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPVGGELIEKVRK 188
DB 136 KYTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPVGGELIEKVRK 195

QY 189 AHQETFPALCQLGKYYTNNSSQORVSLDIDLMDKFSSELSTKCIITKTVFAKQLPGFTT 248
DB 196 AHQETFPALCQLGKYYTNNSSADHRVRLDLGLMDKFSSELATKCIITKTVFAKQLPG 255

QY 249 IADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTLNRTQHNAGFGLTLDLVFAFA 308
DB 256 IADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTLNRTQHNAGFGLTLDLVFAFA 315

QY 309 NOLLPLEMDDAETGLLSAICLCGDRDLQOPDRVDMLOEPLEALKVYVRRRPSRPHM 368
DB 316 GOLLPLEMDDDTETGLLSAICLCGDRDLQOPDRVDMLOEPLEALKVYVRRRPSQPYM 375

QY 369 FPKMLKITDLRSISAKGAERVITLKEIPGSMPLLIQEMLENSEGLDTLSGQGP----- 423
DB 376 FPKMLKITDLRSISAKGAERVITLKEIPGSMPLLIQEMLENSEGLDTLSGQGP----- 435

QY 424 -----GGGRDGGGLPPP 435
DB 436 SSEDEVPGGQGGGLKSP 453
```

```
RESULT 6
US-09-797-727-2
; Sequence 2, Application US/09797727
; Patent No. US20020077457A1
; GENERAL INFORMATION:
; APPLICANT: the Salk Institute for Biological Studies
; APPLICANT: TAKAKU, Fumimaro
; TITLE OF INVENTION: GAMMA RETINOIC ACID RECEPTOR
; FILE REFERENCE: SALK1150-3
; CURRENT APPLICATION NUMBER: US/09/797,727
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 08/486,325
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/100,039
; PRIOR FILING DATE: 1993-07-30
; PRIOR APPLICATION NUMBER: PCT/US90/03564
; PRIOR FILING DATE: 1990-06-22
; PRIOR APPLICATION NUMBER: US 07/370,407
; PRIOR FILING DATE: 1989-06-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Human
; NAME/KEY: misc.feature
; OTHER INFORMATION: Human Retinoic Acid Receptor-gamma (hRAR-gamma)
US-09-797-727-2
```

```

Query Match      68.2%;   Score 1662;   DB 10;   Length 454;
Best Local Similarity 72.8%;   Pred. No. 7.5e-128;
Matches 319;   Conservative 39;   Mismatches 48;   Indels 32;   Gaps 4

QY      30 PEKGLPAPG-PYSNP-----LRPLANG-----SNHSTETQSSSEBIVP 68
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      16 PGSGYPGAGFPFAPFGALRGSPPFEMLSFGRGLGQDPLPKEMASLSVETQSTSSEMPV 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      69 SPPSPPPPLPRYKPCFVCQDKSSGYHYGVSACGCKGFFRRSIQKNMVTCHRDKNCIIN 128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      76 SSPSPPPPRVYKPCFVCNDKSSGYHYGVSCEGCKGFFRRSIQKNMVTCHRDKNCIIN 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      129 KVTNRPCOYKRLQKCFEYGMGSKESVNRDNKKKKEVPKPECSSEYTLPEVGELIEKVRK 188
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      136 KVTNRNCOYKRLQKCFEYGMGSKKEAVNRDNKKKKEVKEGSPDSYELSPQLEELITKYVK 195
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      189 AHQETFFALCOLGKYTTNNSSEQRVSLDLDLWDFSELSKCIKTVEFAKOLPGFTLT 248
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      196 AHQETFFSLCOLGKYTTNSSADHRVQLDLGLWDFSELSATCIIKIVEFAKLPFGTGLS 255
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      249 IADQITLLKAACLDILILRICRTRYPTQDQMTFSDGLTLNRTQMHNAGFPLTDLVFAFA 308
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      256 IADQITLLKAACLDILMLRICRTRYPTQDQMTFSDGLTLNRTQMHNAGFPLTDLVFAFA 315
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      309 NQLPLEMDDAETGLLSAICLICGDRQDLEQPDVDMLOEPLLEALKYVYKRRRPSRPHM 368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      316 GQLPLEMDDETGLLSAICLICGDRMDLEPEKVDKLOEPLLEALRYARRRPSQPYM 375
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      369 FPKMLKITDLRSISAKGAEVITLKMIEPCSMPLIQEMLENSGLDTLSQPG----- 423
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      376 FPKMLKITDLRGISTGAEARITLKMIEPGMPLIREMLENPFEDDSSQPGHPNA 435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      424 -----GGRDGGGLPPP 435
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      436 SSEDEVPGCGKGGLKSP 453
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-10-013-823-3
; Sequence 3, Application US/10013823
; Patent No. US20020116731A1
; GENERAL INFORMATION:
; APPLICANT: Guenther, Catherine
; APPLICANT: Phillips, Russell
; APPLICANT: Allen, Keith D.
; APPLICANT: Zhang, Qin
; APPLICANT: Baribault, Helene
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING RETINOID X
; TITLE OF INVENTION: RECEPTOR INTERACTING PROTEIN GENE DISRUPTIONS
; FILE REFERENCE: R-684
; CURRENT APPLICATION NUMBER: US/10/013,823
; CURRENT FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/254,801
; PRIOR FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: US 60/309,404
; PRIOR FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-013-823-3

```

```

Query Match      23.0%; Score 560; DB 12; Length 461;
Best Local Similarity 30.6%; Pred. No. 8e-38;
Matches 140; Conservative 82; Mismatches 155; Indels 80; Gaps 12;

QY    29 LPEKGLPAPC-pYSTPL-----RTPLWNGSNHSI-----ETQSSSSEIVPSPPSP--- 74
|| || || | || | || | || | || | || | || | || | || | || | || | || |
DB    13 LPGNPPQPQAPSSPIVKEGEPWPGGDDPDVPFGTDEASSACSTDWIPDPBEEPERK 72

```

```

75 ---PLPRI--YPCFCVQDKSSGHHGVACSGCGKFFRRSIOKNMV--YTCHRDKNCI 120
    :|::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
73 RKKGPAPKMLGLHCLRCVGDKASGFHYNLVSCBGCKGFFRRSYVRGGARRIACRGGTQC 130
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
127 INKVTNPCCOYCRLOKCFEVMGSKESVRNDRNKKKEVPKPECSSEYTLT----- 176
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
133 MDAPMRKCOOCLRKCCEAGMREOCVLSEQIRKKKIIRKQQOQSOSQSPVGPGSS 192
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
177 -----PEVG-----ELIEVKRKAHQETFPALCOLGKYT----- 204
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
193 SSASGPGASPGEAGSGSGEGEVQLTAQAELMIQLVAQAQCNKRKSFSDPKVTPW 252
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
205 -----TNNSBQRVSLDIDLWDKSELSKCIITVFBAKOLPGFTLTADQITLLK 257
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
253 PLGADPOSROARQORTA-----HFTELAITSVQEIVDFAKQVPGFLQIGREDQIALLK 305
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
258 AACIDILILRICRYTPEQDTMTFSDGLTLNRQMHNAGFG-PLTDLVFAAFANOLLPLEM 316
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
306 ASTIEIMLETARRYNHETECITFLKDFTYSKDDFRAGLQVEFINPIEFSPAMRRRLGL 365
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
317 DDAETGLLSAICLICGRDLQOPDRDYMDLPLEALKVVYRKRPSPHPFKMLMKI 376
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
366 DDAREYALLIAINTFSDARNVQEPGVEALQOQPYVEALLSYTRIKRPQDLRFPRMLMKL 425
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
377 TDLRSISAGAERVITLKWELPGSMPLIQMLENSE 413
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
426 VSLRTLUSSHVSQVFALRLQ-DKKLPLPULSEIWVDVHE 461
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-09-909-446-2
; Sequence 2, Application US/09909446
; Patent No. US20020052489A1
; GENERAL INFORMATION:
; APPLICANT: ENMARK, EVA
; GUSTAFSSON, JAN
; TITLE OF INVENTION: OR-1 ON ORPHAN RECEPTOR BELONGING TO THE NUCLEAR RECEPTOR FAMILY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/909,446
; FILING DATE: 19-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/7776,844.
; FILING DATE: <Unknown>
; APPLICATION NUMBER: UK 9413536.2
; FILING DATE: 16-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32141
; REFERENCE/DOCKET NUMBER: 00487.04029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```


Db 286 FFARTVQKNAKYVCLANKNCVPDKRRNRRCQYCRFOKCLAVGMVKEVVRTDSLKGRGRGL 345
QY 165 P-KPESESYYLTPEVGEELIEKVRKAHQETFPALCOLG-----KYTTNNSSEQRVSL 215
Db 346 PSKPKSPQDPSPSPVSLISALVRAHVDSPNTPAMTSLDYSRFOANPDYQMSGDDTQHI-- 403
QY 216 DIDLWDFSELSTKCIKIVTFAKQIPGFTTLTADQITLLKAAACLDLILRICTRYTPE 275
Db 404 -----QQFYDILLTGSMEIIRGNAEKIPGADLPKADQDLIFESAFLELVRLAYRNPV 458
QY 276 QDTMTFSDGLTLNRTQMHNAGFPLDVLVAFANQLLPLEMDDAETGLLSAICLCGRQ 335
Db 459 EGKLFPCNGVVLHRLQCVR-GFGEWIDSIVFSSNLQNMNIDISAFSCIAALAMVT-ERH 516
QY 336 DLEQDPDRVDMLOEPLLEALKVYVRKRRP--SRPHMFPKMLKITDLSISAKGAERVITL 393
Db 517 GLKEPKRVEELQNKIVNCLKDHTVFNNGGLNRPNYLSKLLGKLPDLRTCTOGLQRIFYL 576
QY 394 KME--IPGSMPPLIQEMLENSEGLDYL 418
Db 577 KLEDLVP---PPAIDKL-----FLDTL 596

RESULT 15

US-09-853-386-37.

; Sequence 37, Application US/09853386

; Patent No. US20020049151A1

; GENERAL INFORMATION:

; APPLICANT: Murphy, Evelyn

; APPLICANT: Bresnihan, Barry

; APPLICANT: Conneely, Orla

; APPLICANT: Fitzgerald, Oliver

; TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NURR

; FILE REFERENCE: P01972051

; CURRENT APPLICATION NUMBER: US/09/853, 386

; PRIOR FILING DATE: 2001-05-11

; PRIOR APPLICATION NUMBER: US 60/203645

; NUMBER OF SEQ ID NOS: 153

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 37

; LENGTH: 598

; TYPE: PRT

; ORGANISM: HUMAN

US-09-853-386-37

Query Match 21.3%; Score 519.5; DB 10; Length 598;
Best Local Similarity 31.5%; Pred. No. 2.6e-34;
Matches 141; Conservative 79; Mismatches 154; Indels 73; Gaps 16;

QY 24 NRACLLPEKGLPAGPYST-----PLTPLWN-----GSNHSIETQSSSEIIVPSP-- 70

Db 171 SRLSLFSFKQSPGPTPVSSQMRFDGPLHVPN-NPEPAGSHVVDGQTPA----VPNPIR 225

QY 71 -----PSPPP--LPRIYKPCFCODKSSGYHYGSACGCKG 105

Db 226 KPASMGFFPGLQIGHASQLDVTQVPSPPSRGSPSNEGLCAVCGDNACQHYGVRTCEGCKG 285

QY 106 FFRSIOKNMVTYCHRDKNCIINKVTRNQCQYCRKQCFEVMGSKESVRNDRNK-KKKEV 164

Db 286 FFRTVQKNAKYVCLANKNCVPDKRRNRRCQYCRFOKCLAVGMVKEVVRTDSLKGRGRGL 345

QY 165 P-KPESESYYLTPEVGEELIEKVRKAHQETFPALCOLG-----KYTTNNSSEQRVSL 215

Db 346 PSKPKSPQDPSPSPVSLISALVRAHVDSPNTPAMTSLDYSRFOANPDYQMSGDDTQHI-- 403

QY 216 DIDLWDFSELSTKCIKIVTFAKQIPGFTTLTADQITLLKAAACLDLILRICTRYTPE 275

Db 404 -----QQFYDILLTGSMEIIRGNAEKIPGADLPKADQDLIFESAFLELVRLAYRNPV 458

QY 276 QDTMTFSDGLTLNRTQMHNAGFPLDVLVAFANQLLPLEMDDAETGLLSAICLCGRQ 335

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:00:38 ; Search time 2120.41 Seconds
(without alignments)
13725.086 Million cell updates/sec

Title: US-09-691-220-3_COPY_1_1000
Perfect score: 1000
Sequence: 1 gtccttggttagcatgtaca.....acatgcgcgcctctccctctc 1000

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	973.6	97.4	16913	9	HSA297538	AJ297538 Homo sapi
C 2	973.6	97.4	138999	9	AC090426	AC090426 Homo sapi
C 3	973.6	97.4	173441	2	AC080112	AC080112 Homo sapi
C 4	962.6	96.3	188574	2	AC131063	AC131063 Homo sapi
C 5	962.6	96.3	190309	2	AC126392	AC126392 Homo sapi
C 6	956.8	95.7	158766	2	AC015851	AC015851 Homo sapi
7	875.6	87.6	1833	9	AF283809	AF283809 Homo sapi
8	263.2	26.3	231248	2	AL591067	AL591067 Mus muscu
	259	25.9	1330	10	MUSARA21	M80781 Mouse reti
C 10	242.8	24.3	13508	6	AX344791	AX344791 Sequence
C 11	228	22.8	2008	9	AK098172	AK098172 Homo sapi
C 12	143.8	14.4	13508	6	AX344790	AX344790 Sequence
C 13	127	12.7	162393	2	AC111746	AC111746 Rattus no
C 14	73.8	7.4	972	11	PM12A12G	AL684288 Penicill
C 15	72.6	7.3	7218	6	I66494	I66494 Sequence 14
C 16	69.4	6.9	90352	2	AC099278	AC099278 Rattus no
C 17	69.2	6.9	207420	2	AC078884	AC078884 Mus muscu
C 18	68.8	6.9	146395	2	AC127918	AC127918 Rattus no
C 19	67.4	6.7	300695	2	AC079431	AC079431 Mus muscu
C 20	67	6.7	94029	2	AC117309	AC117309 Rattus no
C 21	66.6	6.7	152261	2	AC121010	AC121010 Rattus no
C 22	66.4	6.6	112299	2	AC125732	AC125732 Rattus no
C 23	65.4	6.5	130626	2	AC113687	AC113687 Rattus no
C 24	65	6.5	177801	2	AC114391	AC114391 Rattus no
C 25	65	6.5	186882	2	AC095077	AC095077 Rattus no
C 26	65	6.5	227009	2	AL845264	AL845264 Mus muscu
C 27	64.6	6.5	208978	2	AC126593	AC126593 Rattus no
C 28	64.4	6.4	125020	9	AF429315	AF429315 Homo sapi
C 29	64.4	6.4	145592	2	AC119341	AC119341 Rattus no
C 30	63.8	6.4	83367	2	AC115654	AC115654 Rattus no
C 31	63.8	6.4	141905	2	AC110101	AC110101 Rattus no
C 32	63.6	6.4	119989	2	AC119472	AC119472 Rattus no
C 33	63.6	6.4	151283	2	AC107540	AC107540 Rattus no
C 34	63.4	6.3	152895	2	AC105671	AC105671 Rattus no
C 35	63.2	6.3	76734	2	AC023218	AC023218 Homo sapi
C 36	63.2	6.3	212355	2	AC095539	AC095539 Rattus no
C 37	62.6	6.3	148481	2	AC127076	AC127076 Rattus no
C 38	62.6	6.3	174605	2	AC121227	AC121227 Rattus no
C 39	62.4	6.2	42748	2	AC105855	AC105855 Rattus no
C 40	62.4	6.2	184402	2	AC127041	AC127041 Rattus no
C 41	62.2	6.2	152732	2	AC074204	AC074204 Mus muscu
C 42	62.2	6.2	182269	2	AC107416	AC107416 Rattus no
C 43	62	6.2	158216	2	AC129233	AC129233 Rattus no
C 44	61.8	6.2	125020	9	AF429315	AF429315 Homo sapi
C 45	61.8	6.2	175440	2	AC126197	AC126197 Rattus no

ALIGNMENTS

RESULT 1
HSA297538
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
REFERENCE

HSA297538
Homo sapiens partial RARA gene, intron 2.
AJ297538
AJ297538.1 GI:12054223
Rara gene; retinoic acid receptor alpha.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 16913)
Cross,N.C.P. and Reiter,A.
Unpublished
2 (bases 1 to 16913)

linear PRI 06-JAN-2001

AUTHORS Cross,N.C.P.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-2000) Cross N.C.P., Department of Haematology,
Imperial College School of Medicine, Hammersmith Hospital, London,
W12 0NN, UNITED KINGDOM

FEATURES Location/Qualifiers
source 1.16913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q12"

gene 1.16913
/gene="RARA"
intron 1.16913
/gene="RARA"

repeat_region 1638..1947
/rpt_family="AluSg"
repeat_region 1949..2235
/rpt_family="AluSx"
repeat_region 2236..2536
/rpt_family="AluJo"
repeat_region 4008..4315
/rpt_family="AluSx"
repeat_region 4327..4619
/rpt_family="AluJb"
repeat_region 4669..4975
/rpt_family="AluSx"
repeat_region 4982..5009
/note="(TTTA)n"
repeat_region 5011..5287
/rpt_family="AluY"
repeat_region 5287..5597
/rpt_family="AluSc"
repeat_region 5666..5776
/rpt_family="MIR"
repeat_region 7090..7148
/rpt_family="MIR"
repeat_region 7159..7326
/rpt_family="AluSg/x"
repeat_region 7331..7641
/rpt_family="AluJb"
repeat_region 8518..8659
/rpt_family="L2"
repeat_region 8853..9194
/rpt_family="L2"
repeat_region 10145..10198
/note="(CGGG)n"

BASE COUNT 3169 a 4730 c 4922 g 4092 t
ORIGIN

Query Match 97.4%; Score 973.6; DB 9; Length 16913;

Best Local Similarity 99.0%; Pred. No. 1.5e-188;
Matches 990; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 1 GTCCTGGGTAGCATGATTCATCCCTTCCTTTATATATGCGGGTAAATAGATAC 60
Db 9621 GTCCTGGGTAGCATGATTCATCCCTTTAGAGAGTGGGGTAAATAGATAC 9680
Qy 61 CCCCTCTCCAGGGTATCCCTCTTTTAGGGACCTACCCAAAGCTAGCCCTTTCTCCA 120
Db 9681 CCCCTCTCCAGGGTATCCCTCTTTTAGGGACCTACCCAAAGCTAGCCCTTTCTCCA 9740
Qy 121 GTGAACCTGATCCCGAGGGCTTCTAGATGAAGTAGTCCACTGGAGGCACCACTCT 180
Db 9741 GTGAACCTGATCCCGAGGGCTTCTAGATGAAGTAGTCCACTGGAGGCACCACTCT 9800
Qy 181 TCCCTTTATCTCCAGAGCTGGACAGTGCACAGGGCCGGTACTGTTCCCGAGCTAG 240
Db 9801 TCCCTTTATCTCCAGAGCTGGACAGTGCACAGGGCCGGTACTGTTCCCGAGCTAG 9860
Qy 241 GAGACACTTGGGGGGCTTTGCTCGCCGGAAGCACAGAGCTGGGGAGGGGCC 300
|||||

Db 9861 GAGACACTTGGCGGGGCTTTGCTCCCGGAAGCACAGCAGCGTGGGAGGAGGCC 9920
Qy 301 CCTCTGCTGTGTTTGTGCCAACAGCACCAGCGCTGCCGCGTGGGTTCCGGCGGCCGGA 360
Db 9921 CCTCTGCTGTGTTTGTGCCAACAGCACCAGCGCTGCCGCGTGGGTTCCGGCGGCCGGA 9980
Qy 361 GTCACACATGATGTACAGACATGACAAAGCGGTGTCTCATTCGACACACAGCGTCCG 420
Db 9981 GTCACACATGATGTACAGACATGACAAAGCGGTGTCTCATTCGACACACAGCGTCCG 10040
Qy 421 AGCTGCACAATGTACACCCGGGTGCCAAACACTTGGCCCCGCGCAGCCGCGCTACGC 480
Db 10041 AGCTGCACAATGTACACCCGGGTGCCAAACACTTGGCCCCGCGCAGCCGCGCTACGC 10100
Qy 481 CTCTGCGCGCGCTCTCCGGGTCTCCGGGGAGGTGGCCCGGTTCCGCCGGGAGGGGGC 540
Db 10101 CTCTGCGCGCGCTCTCCGGGTCTCCGGGGAGGTGGCCCGGTTCCGCCGGGAGGGGGC 10160
Qy 541 TGGCGGGCGAGCCCGCGGGCTGGCGAGCGGGTGTGTACCGGCGCAGCGGTGGGTG 600
Db 10161 TGGCGGGCGAGCCCGCGGGCTGGCGAGCGGGTGTGTACCGGCGCAGCGGTGGGTG 10220
Qy 601 GGTCACTCGGAGGTGAGCGCCCGCAGCGAGTTACAGCAGAGTTACAGCCGCAATGCATT 660
Db 10221 GGTCACTCGGAGGTGAGCGCCCGCAGCGAGTTACAGCAGAGTTACAGCCGCAATGCATT 10280
Qy 661 AGGCAATGAGCGCGCGCTGGTGGGGTGTGTGTAAAGGGAGGACACCGGGACCAACC 720
Db 10281 AGGCAATGAGCGCGCGCTGGTGGGGTGTGTGTAAAGGGAGGACACCGGGACCAACC 10340
Qy 721 CCCCTCTTCCCGCGCCACACCTCTCCACACAGCGTTCGCTCGCGCAGGACTGACCA 780
Db 10341 CCCCTCTTCCCGCGCCACACCTCTCCACACAGCGTTCGCTCGCGCAGGACTGACCA 10400
Qy 781 AACCTTGGGGAGCGCTGGAGCGGGAACGTGTATACAGGGAGAGACGCCGCCCTCTTCC 840
Db 10401 AACCTTGGGGAGCGCTGGAGCGGGAACGTGTATACAGGGAGAGACGCCGCCCTCTTCC 10459
Qy 841 GTCTTGTCTCCCTCGCAGCGCCCTCTCTCCCTGACTCTGCTGCTCTCTCTCTCTCTCTCT 900
Db 10460 GTCTTGTCTCCCTCGCAGCGCCCTCTCTCTCTGACTCTGCTGCTCTCTCTCTCTCTCT 10519
Qy 901 TACTCTCATCTGGAGCCTTTCCCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db 10520 TACTCTCATCTGGAGCCTTTCCCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10579
Qy 961 CTGCCCCCACTTGCCTGTCCACATGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1000
Db 10580 CTGCCCCCACTTGCCTGTCCACATGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10619

RESULT 2

AC090426/c

LOCUS

DEFINITION

AC090426

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

12032336

138999 bp

DNA

linear

PRI 29-MAY-2002

map 17q21.1, complete

sequence.

AC090426

GI:12965315

HTG.

Homo sapiens

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 138999)

Gu.B., Xiong.H., Zhou.Y., Chen.B., Lu.L., Zhong.M., Yin.H.,

Huang.W., Ren.S., Chen.S.F., Chen.Z. and Fu.G.

Variant-type PMR-RAR(alpha) fusion transcript in acute

promyelocytic leukemia: use of a cryptic coding sequence from

intron 2 of the RAR(alpha) gene and identification of a new

clinical subtype resistant to retinoic acid therapy

Proc. Natl. Acad. Sci. U.S.A. 99 (11), 7640-7645 (2002)

22028997

12032336

```

REFERENCE 2 (bases 1 to 138999)
AUTHORS Gu,B., Xiong,H., Zhou,Y., Chen,B., Lu,L., Zhong,M., Yin,H.,
Huang,W., Ren,S., Chen,S.F., Chen,Z. and Fu,G.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2001) Chinese National Human Genome Center at
Shanghai, Shanghai, Shanghai 201203, P.R.China
COMMENT -----Genome Center-----
Center:Chinese National Human Genome Center at Shanghai
Center code:CHGC Website: http://www.chgc.sh.cn Contact:
fugang@chgc.sh.cn
FEATURES
source Location/Qualifiers
1..138999 /organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q21.1"
/clone="205ml7"
repeat_region 387..483 /rpt_family="SINE/MIR"
/rpt_unit="MIR"
repeat_region 725..1032 /rpt_family="SINE/Alu"
/rpt_unit="AluJb"
repeat_region 1033..1198 /rpt_family="SINE/Alu"
repeat_region 1217..1514 /rpt_family="SINE/Alu"
/rpt_unit="AluSp"
repeat_region complement(1534..1754) /rpt_family="SINE/MIR"
/rpt_unit="MIR"
repeat_region 2291..2366 /rpt_family="SINE/MIR"
/rpt_unit="MIR"
repeat_region complement(2607..2665) /rpt_family="SINE/MIR"
/rpt_unit="MIR"
repeat_region 3651..3905 /rpt_family="LINE/L1"
/rpt_unit="L1ME2"
repeat_region 3977..4029 /rpt_family="LINE/L1"
/rpt_unit="L1ME2"
repeat_region 4030..4213 /rpt_family="LTR/MaLR"
/rpt_unit="MSTD"
repeat_region 4214..4509 /rpt_family="SINE/Alu"
/rpt_unit="AluJb"
repeat_region 4514..4811 /rpt_family="SINE/Alu"
/rpt_unit="AluSx"
repeat_region 4812..5018 /rpt_family="LTR/MaLR"
/rpt_unit="MSTD"
repeat_region 5019..5102 /rpt_family="LINE/L1"
/rpt_unit="L1ME2"
repeat_region 5236..5443 /rpt_family="SINE/MIR"
/rpt_unit="MIR"
repeat_region 5467..5699 /rpt_family="SINE/MIR"
/rpt_unit="MIR"
repeat_region 6547..6708 /rpt_family="SINE/MIR"
repeat_region 6827..6917 /rpt_family="Low_complexity"
/rpt_unit="GA-rich"
repeat_region 7733..7779 /rpt_family="SINE/MIR"
/rpt_unit="MIR"
8268..8411 /rpt_family="Simple_repeat"
/rpt_unit="(TGGG)n
9224..9532 /rpt_family="SINE/Alu"
/rpt_unit="AluY"
9546..9639 /rpt_family="Low_complexity"
/rpt_unit="Cr-rich"
10025..10426 /rpt_family="LTR/Retroviral"
/rpt_unit="MLT2A"
10433..10471 /rpt_family="Simple_repeat"
/rpt_unit="(TA)n
10472..10527 /rpt_family="LTR/Retroviral"
/rpt_unit="MLT2A"
complement(10675..10962) /rpt_family="SINE/Alu"
/rpt_unit="AluY"
11013..11321 /rpt_family="SINE/Alu"
/rpt_unit="AluY"
complement(11840..12192) /rpt_family="LTR/MaLR"
/rpt_unit="MSTD"
12729..12835 /rpt_family="SINE/Alu"
/rpt_unit="FLAM_A"
complement(12841..12891) /rpt_family="LINE/L2"
/rpt_unit="L2"
complement(14228..14548) /rpt_family="SINE/Alu"
/rpt_unit="AluY"
complement(14618..14701) /rpt_family="SINE/MIR"
/rpt_unit="MIR"
14953..15255 /rpt_family="SINE/Alu"
/rpt_unit="AluJo"
15311..15473 /rpt_family="Simple_repeat"
/rpt_unit="(TA)n
15669..15981 /rpt_family="SINE/Alu"
/rpt_unit="AluSx"
16316..16353 /rpt_family="Simple_repeat"
/rpt_unit="(TAAA)n
complement(16863..16910) /rpt_family="SINE/MIR"
/rpt_unit="MIR"
17026..17324 /rpt_family="SINE/Alu"
/rpt_unit="AluSg"
17337..17618 /rpt_family="SINE/Alu"
/rpt_unit="AluSx"
17619..18024 /rpt_family="LTR"
/rpt_unit="LTR54"
complement(18026..18086) /rpt_family="SINE/MIR"
/rpt_unit="MIR"
complement(18292..18607) /rpt_family="SINE/Alu"
/rpt_unit="AluYb8"
complement(18948..19264) /rpt_family="SINE/Alu"
/rpt_unit="AluSx"
complement(20012..20168)

```


JOURNAL

REFERENCE
AUTHORS

Submitted (24-SEP-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 173441)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., O'Connor,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 14, 2002 this sequence version replaced gi:22123315.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L11030
Center clone name: 2267_D_19

TITLE

JOURNAL

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 2128: contig of 2128 bp in length
* 2129 2228: gap of 100 bp
* 2229 157931: contig of 155703 bp in length
* 157932 158031: gap of 100 bp
* 158032 173441: contig of 15410 bp in length.
Location/Qualifiers
1. 173441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="CTD-2267D19"
/clone_lib="CITD1 Human BAC"
46772 a 43344 c 40918 g 42114 t 293 others

FEATURES
source

BASE COUNT
ORIGIN

Query Match 97.4%; Score 973.6; DB 2; Length 173441;
Best Local Similarity 99.0%; Pred. NO. 1.2e-186;
Matches 990; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
Qy 1 GTCCTGGGTAGCATGTCATCCCTCCCTTTATATATATGGGGTATAGATAC 60
|||||
Db 82724 GTCCTGGGTAGCATGTCATCCCTCCCTTTATAGAGTGGGGTATAGATAC 82665
|||||
Qy 61 CCCCTCCCTCCAGGGGTATCCCTCTCTTAGGGACCTACCCCAAGCTAGGCCCTTCTTCCA 120
|||||
Db 82664 CCCCTCCCTCCAGGGGTATCCCTCTCTTAGGGACCTACCCCAAGCTAGGCCCTTCTTCCA 82605
|||||

Qy 121 GTCAACAGTGCATCCGAGGGGCTTCTAGGATGAAGTAGTCCACTGGAAGCACACAGCTCT 180
|||||
Db 82604 GTGAACAGTGCATCCGAGGGGCTTCTAGGATGAAGTAGTCCACTGGAAGCACACAGCTCT 82545
|||||
Qy 181 TCCCTTTATCTCTCCAGAGCTGGACAGTGCACACAGGGGCGGTACTGTTTCCCCAGCTAG 240
|||||
Db 82544 TCCCTTTATCTCTCCAGAGCTGGACAGTGCACACAGGGGCGGTACTGTTTCCCCAGCTAG 82485
|||||
Qy 241 GAGACACCTTGGGGGGGCTTGTCTCCGCGAAGACAGACAGCGTGGGGAGGAGGGCCC 300
|||||
Db 82484 GAGACACCTTGGGGGGGCTTGTCTCCGCGAAGACAGACAGCGTGGGGAGGAGGGCCC 82425
|||||
Qy 301 CCTCTGCGCTGTCTTGTGCGCAACAGCACCGCGCTCCGCGTCCGGTTCGCGCGCGCGA 360
|||||
Db 82424 CCTCTGCGCTGTCTTGTGCGCAACAGCACCGCGCTCCGCGTCCGGTTCGCGCGCGCGA 82365
|||||
Qy 361 GTCACACATGATGTACACAGCAATGACACAAGCGGTGTCTCATTCGACACAGCGTCCG 420
|||||
Db 82364 GTCACACATGATGTACACAGCAATGACACAAGCGGTGTCTCATTCGACACAGCGTCCG 82305
|||||
Qy 421 AGCTGCACATGTACACACCGGGGTGCGCAACACTTGGCCCCGCGACCGCGCCCTTACGC 480
|||||
Db 82304 AGCTGCACATGTACACACCGGGGTGCGCAACACTTGGCCCCGCGACCGCGCCCTTACGC 82245
|||||
Qy 481 CTCCTGCCCGCTCTCCCGGTCTCCGGGGAGGTGCGCGGTTCGCGCGCGAGGGGGC 540
|||||
Db 82244 CTCCTGCCCGCTCTCCCGGTCTCCGGGGAGGTGCGCGGTTCGCGCGCGAGGGGGC 82185
|||||
Qy 541 TGGCGGGCGAGCGCGCGGGCTGGCGAGCGGGTGATGTACACGGGCGACGCGTGGGGTG 600
|||||
Db 82184 TGGCGGGCGAGCGCGCGGGCTGGCGAGCGGGTGATGTACACGGGCGAGCGTGGGGTG 82125
|||||
Qy 601 GGTACTCGGAGGTGAGGCGCGCGCGAGGTTCAGCGAGAGTTGAGCGCATTTGCATT 660
|||||
Db 82124 GGTACTCGGAGGTGAGGCGCGCGCGAGGTTCAGCGAGAGTTGAGCGCATTTGCATT 82065
|||||
Qy 661 AGGCAATGAGCGCGCGCTGGGTGGGGTGTTAGGGGAGGACACCGGGACACCC 720
|||||
Db 82064 AGGCAATGAGCGCGCGCTGGGTGGGGTGTTAGGGGAGGACACCGGGACACCC 82005
|||||
Qy 721 CCCCTCTTCCCGCGCGCGCGCTCCCTCCACACAGCGCTTCGCTCGGCGAGGACTGACCA 780
|||||
Db 82004 CCCCTCTTCCCGCGCGCGCGCTCCCTCCACACAGCGCTTCGCTCGGCGAGGACTGACCA 81945
|||||
Qy 781 AACCTTGGGGAGCTTGGGAGCGCGGAACTGGTACAGGGGAGGACGCGCGCCCTCTTTC 840
|||||
Db 81944 AACCTTGGGGAGCTTGGGAGCGCGGAACTGGTACAA- GGGAGGACGCGCGCGCTCTTC 81886
|||||
Qy 841 GTCCTTGTCCCTCGCAGCGCGCTCCCTCTCTCTGTTACTCGCGCTCCTCTGTACTCTGTG 900
|||||
Db 81885 GTCCTTGTCCCTCGCAGCGCGCTCCCTCTCTCTGTTACTCGCGCTCCTCTGTACTCTGTG 81826
|||||
Qy 901 TACTCTCATCTGGAGCTTTCCCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 960
|||||
Db 81825 TACTCTCATCTGGAGCTTTCCCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 81766
|||||
Qy 961 CTGCCCCCACTGCTGTCACATGCGCGCTCTCCCTCTC 1000
|||||
Db 81765 CTGCCCCCACTGCTGTCACATGCGCGCTCTCTCTCTC 81726
|||||

RESULT 4
AC131063/c
LOCUS
DEFINITION Homo sapiens chromosome 17 clone RP11-1029L16 map 17, WORKING DRAFT
SEQUENCE, 6 unordered pieces.
ACCESSION AC131063
VERSION AC131063.1 GI:22267816
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.


```
QY 720 CCCCTCTTCCCGCCGACACCTCTCTCCACAGGCTTCGTCGCCGACGAGTGCAC 779
|||||
Db 165606 CCCCTCTTCCCGCCGACACCTCTCTCCACAGGCTTCGTCGCCGACGAGTGCAC 165547
|||||
QY 780 AACTCTGGGGAGCTGGGAGCCGGAAGTGTACAGGAGGAGGAGCGCCGCTCTTC 839
|||||
Db 165546 AACTCTGGGGAGCTGGGAGCCGGAAGTGTACAA-GGGAGGAGCGCCGCTCTTC 165488
|||||
QY 840 GTCCTCTGTCCTCGACGCCCTCTCTCCCTGTACTGCGGTCCCTCTGTACTCTGT 899
|||||
Db 165487 GTCCTCTGTCCTCGACGCCCTCTCTCCCTGTACTGCGGTCCCTCTGTACTCTGT 165428
|||||
QY 900 GTACTCTCATCTGAGCGCTTTCCTCCCTGCTCTCTCTCTCTCTCTCTCTCTCC 959
|||||
Db 165427 GTACTCTCATCTGAGCGCTTTCCTCCCTGCTCTCTCTCTCTCTCTCTCTCTCC 165368
|||||
QY 960 GCTGCCCCACCTGCTGTCACATGCGCGCTCTCCCTC 1000
|||||
Db 165367 GCTGCCCCACCTGCTGTCACATGCGCGCTCTCCCTC 165327
|||||

RESULT 5
AC126392/c
LOCUS
DEFINITION Homo sapiens chromosome 17 clone RP11-1029F16 map 17, WORKING DRAFT
SEQUENCE, 11 ordered pieces.
ACCESSION AC126392
VERSION AC126392.2 GI:221233095
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 190309)
Barra, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
3 (bases 1 to 190309)
Barra, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
3 (bases 1 to 190309)
Barra, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT
On Aug 6, 2002 this sequence version replaced gi:21699357.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27553
Center clone name: 1029_F16
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 187521 bases at least Q40
Consensus quality: 188193 bases at least Q30
Consensus quality: 188537 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 189309; sum-of-contigs
Quality coverage: 15.5 in Q20 bases; agarose-fp
Quality coverage: 14.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 30513: contig of 30513 bp in length
* 30514 30613: gap of 100 bp
* 30614 31764: contig of 1151 bp in length
* 31765 31864: gap of 100 bp
* 31865 33693: contig of 1829 bp in length
* 33694 33793: gap of 100 bp
* 33794 35405: contig of 1612 bp in length
* 35406 35505: gap of 100 bp
* 35506 37691: contig of 2186 bp in length
* 37692 37791: gap of 100 bp
* 37792 51370: contig of 13579 bp in length
* 51371 51470: gap of 100 bp
* 51471 66014: contig of 14544 bp in length
* 66015 66114: gap of 100 bp
* 66115 108183: contig of 42069 bp in length
* 108184 108283: gap of 100 bp
* 108284 130510: contig of 22227 bp in length
* 130511 130610: gap of 100 bp
* 130611 166740: contig of 36130 bp in length
* 166741 166840: gap of 100 bp
* 166841 190309: contig of 23469 bp in length.
FEATURES
Location/Qualifiers
1..190309
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-1029F16"
/clone_lib="RPC1-11 Human Male BAC"
1..30513
misc_feature
/feature="assembly_fragment"
```

[illegible]

```

----- Summary Statistics
Sequencing vector: M13; M77615: 41% of reads
Sequencing vector: Plasmid; L08752: 4% of reads
Sequencing vector: Plasmid; n/a; 56% of reads
Chemistry: Dye-primer-amersham; 13% of reads
Chemistry: Dye-terminator Big Dye; 88% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 153948 bases at least Q40
Consensus quality: 156118 bases at least Q30
Consensus quality: 157160 bases at least Q20
Insert size: 63000; agarose-ftp
Insert size: 157866; S.

* NOTE: this is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

Query Match	95.7%	Score	956.8	DB 2	Length	158766			
Best Local Similarity	98.5%	Pred. No.	3.2e-185						
Matches	987	Conservative	0	Mismatches	12	Indels	3	Gaps	

QY	1	GTCTCTGGGTAGCATGTACATTTCATCCCTTCCTTTTATATATATGGGGTAAATAGGATAC	60
DB	72122	GTCTCTGGGTAGCATGTACATTTCATCCCTTCCTTTTATAGAGTGGGGTAAATAGGATAC	72181
QY	61	CCCTCTCTCCAGGGGTATCCCTCTTTCTTAGGAGACCTACCCAAGCTAGGCCCTTTCTTCCA	120
DB	72182	CCCTCTCTCCAGGGGTATCCCTCTTTCTTAGGAGACCTACCCAAGCTAGGCCCTTTCTTCCA	72241
QY	121	GTGAACGTGTGATCCCGAGGGCTTCTTAGTAGAAGTAGTCCACTGTGAAGGCACAGCTCT	180
DB	72242	GTGAACGTGTGATCCCGAGGGCTTCTAGTAGAAGTAGTCCACTGTGAAGGCACAGCTCT	72301
QY	181	TCCTTTTATCTCTCCAGAGCTGGACGTGACACAGGGGCGGTACTGTTTCCCCAGCTAG	240
DB	72302	TCCTTTTATCTCTCCAGAGCTGGACGTGACACAGGGGCGGTACTGTTTCCCCAGCTAG	72361
QY	241	GAGACACCTTGGGGGGGCTTTGCTCGCCGGAAGCAGCAGAGCGTGGGAGGAGGCC	300
DB	72362	GAGACACCTTGGGGGGGCTTTGCTCGCCGGAAGCAGCAGAGCGTGGGAGGAGGCC	72421
QY	301	CCTCTGCGCTGTGTTGTGCCAACAGCACCCGCGTGCCTGCGGTTCGGCGCGCCGGA	360
DB	72422	CCTCTGCGCTGTGTTGTGCCAACAGCACCCGCGTGCCTGCGGTTCGGCGCGCCGGA	72481
QY	361	GTCCACATGATGTCCAGACAATGACACAGCCGGTGTCTCATTCGCACACAGCTCCG	420
DB	72482	GTCCACATGATGTCCAGACAATGACACAGCCGGTGTCTCATTCGCACACAGCTCCG	72541
QY	421	AGCTGCACAATGTCCACACCCGGGTGCCAAACACTTGGCCCGCGCGACCCGGCCCTACGC	480
DB	72542	AGCTGCACAATGTCCACACCCGGGTGCCAAACACTTGGCCCGCGCGACCCGGCCCTACGC	72601
QY	481	CTCTTGCCGCGCTCTCGCGCTCTCCGGGGA--GTTGGCCCGGTTCGGCGGCGCAGGG	538
DB	72602	CTCTTGCCGCGCTCTCGCGCTCTCCGGGGAAGGTGGCCCGGTTCGGCGGCGCAGGG	72661
QY	539	GCTGGGGGCGAGCCCGCGGGCGGCTGGCGAGCGGTGATGTACGGGACGCGTGGG	598
DB	72662	GCTGGGGGCGAGCCCGCGGGCGGCTGGCGAGCGGTGATGTACGGGACGCGTGGG	72721
QY	599	TGGGTCACTCGGAGGTGAGGCGCGCCAGGCGAGTTTCAGCGAGAGTTTCAGCCGATTTGCA	658
DB	72722	TGGGTCACTCGGAGGTGAGGCGCGCCAGGCGAGTTTCAGCGAGAGTTTCAGCCGATTTGCA	72781
QY	659	TTAGGCAATGAGCCCGCGCTGGGTGGGGTGTGTGTTAAGGGGAGGACACCGGGACCA	718
DB	72782	TTAGGCAATGAGCCCGCGCTGGGTGGGGTGTGTGTTAAGGGGAGGACACCGGGACCA	72841
QY	719	CCCCCTCTTCCCGCCGCCACCTCTCTCACACAGCGTTTCGCTCGGCGCAGGACTGAC	778
DB	72842	CCCCCTCTTCCCGCCGCCACCTCTCTCACACAGCGTTTCGCTCGGCGCAGGACTGAC	72901
QY	779	CAAACTTGGGGAGCTGGGAGCCGGAATCTGTTACAAAGGGGAGGAGCCCGCCCTCTT	838
DB	72902	CAAACTTGGGGAGCTGGGAGCCGGAATCTGTTACAAAGGGGAGGAGCCCGCCCTCTT	72960
QY	839	CCGTCCTTGTCCCTTCGACGCCCTCTCTCTCCCTGTACTCGGGCTCCCTGTACTCTG	898
DB	72961	CCGTCCTTGTCCCTTCGACGCCCTCTCTCTCCCTGTACTCGGGCTCCCTGTACTCTG	73020
QY	899	TGTACTCTCATCTGGAGCCTTTCCCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	958
DB	73021	TGTACTCTCATCTGGAGCCTTTCCCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	73080
QY	959	GGCTGCCCCACTTGCTGTGCACATGCCGCTCTCCCTCTC	1000
DB	73081	GGCTGCCCCACTTGCTGTGCACATGCCGCTCTCCCTCTC	73122

[illegible]

Center project name: bm333D2
----- Summary Statistics -----
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator; 1% of reads
Chemistry: Dye-terminator; 1% of reads
Consensus quality: 231046 bases at least Q40
Consensus quality: 231133 bases at least Q30
Consensus quality: 231138 bases at least Q20
Insert size: 231148; sum-of-contigs
Insert size: 232554; 2.5% error; agarose-fp
Quality coverage: 11.23x in Q20 bases; sum-of-contigs Quality
coverage: 11.16x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 134201: contig of 134201 bp in length
* 134202 134301: gap of 100 bp
* 134302 231248: contig of 96947 bp in length.

FEATURES

source	Location/Qualifiers
1. 231248	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/chromosome="11"
	/clone_lib="RPC1-23"
	/clone="RP23-333D2"
misc_feature	1. 134201
	/note="assembly fragment: 06314"
misc_feature	134302..231248
	/note="assembly fragment: 07956"
BASE COUNT	55530 a 56205 c 57206 g 62207 t 100 others
ORIGIN	
Query Match	26.3%; Score 263.2; DB 2; Length 231248;
Best Local Similarity	65.1%; Pred. No. 6.4e-44;
Matches	635; Conservative 0; Mismatches 248; Indels 93; Gaps 13;
QY 34	CTTTTATATATGGGGTAATAGGATACCC-CCTCTCCAGGGGATCCCTCTTCCTTCAGG 92
Db 137655	CTTTTAAAGAGAGATAGGATACCCCTGCTCTCTCTGGGTATCTCTCTTTCTTAGG 137714
QY 93	GACCTACCCAGCTAGCGCTTCTTCCAGTGAACGTGCATCCCGAGGGCTTCAGGATG 152
Db 137715	GACCTACTAACCCAGCCCTTCTTCCACTCAACAGCTCGCTTAG-----GGAGG 137765
QY 153	AAGTAGTCCACTGAAGCCAGCCAGCTCTTCTTTATCTCTCCAGAGCTGGACAGTGCAC 212
Db 137766	AAGTGCTCAACTTAAGTC-----TCTTTTATCTGTCTCTGGCT----- 137805
QY 213	CAGGGGCGGTACTGTTCCCACTAGGAGACACCTTGGCGGGGGCTTGTCTCGCGGA 272
Db 137806	-----GGCCCATCTCTGTTCCCGGCTGGGAGACACCTGGCGGAGCTTGTCTCGCGGA 137861
QY 273	AGCACGAGAGCTGGGGAGGAGGGCCCTCTGCTCTGTTGTTGGCAACAGC----- 326
Db 137862	AGCACGAGAGTGGGAGGAGAGCTCCCC-----GCTTGTGTGTGTGTGTAAACAGCTCAGTT 137918
QY 327	-----ACCCGCGCTGCGCGCTGGGTTCCGGCGCGGAGTCACACATGATGTACA 378
Db 137919	CTGCCGTGTCTTATCTCCGGCGGTGAGCGGCGGAGTCTCATATGATGTACA 137978
QY 379	GACAATGACAAAGCCGG-TGTCTCAATTCGACACAGCGTCCGAGTGCACAAATGTACA 437
Db 137979	GACAGTGACAAAGCTGGATGTCTCAATTCGACACAGCGCCCAAGCTGCACAAATGTACA 138038
QY 438	CCCGGGTGCCAAA-----CACTTGGCCCGCGGAGCCCGGCGCTACGGCTC 483

Db 138039 CCCGGGACACCCAGCATGCTGTCATTCGTGGTTCTAACGCTGGGCCACCTGCAATTGGCC 138098

QY 484 CTGCGCGCGCTCTCCGCGTCTCCGGGGAGGTGGCCCGGTTCGGCCGGGAGGGGCTGG 543

Db 138099 CTGTCCCTACTCCCGCTGTGTTCGGAGTTTGGGGCGGACAGTCCGAGATGGGGCCAG 138158

QY 544 CGGGGAGCCCGCGGCGGCTGCGGAGCGGTGATGTCAGGGGAGCGGTGGGTGGGT 603

Db 138159 TGGCGAGCCCGGAGGGCGGTGCGAGCTGATGTCAGCGGAGCGGTGGGTGGGT 138218

QY 604 CACTCGAGGTGAGCGCGCGGCGGAGTTCAGCGAGAGTTCAGCGGATTCGATTCATTAGG 663

Db 138219 CAGTCGAGGTGAGCGCGCGGCGGAGTTCAGCAAGAGTTCAGCGGATTCGATTCAGG 138278

QY 664 CAATAGGCGCGGCTGGGTGGGTGGGTGTTG-TTAAGGGAGGACACCGGGACACCC 722

Db 138279 CAATAGAGCGGGCTTTGGGGGGGGCTGTTTAAAGGGAGGACACCGGGACACCC 138338

QY 723 CTTCTCCCCCGCCACCACTCTCCACCAAGCTTCTCGCTCGGCGGAGTGCACA-A 781

Db 138339 -----ACCTCCCGCGCTGCTATCTTCCTGCGAGGGATTGACCA 138379

QY 782 ACCTTGGGGAGCCCTGGAGCGGAACTGTATACAGGGAGGAGCGCCCGCTCTTCG 841

Db 138380 AGCACTAGGAGCTGGAGCGGAACTGTG--GATGGGAGAGTGCCTCTACTCTTCG 138437

QY 842 TCTTGTCTCCCGAGCCCTCTCTCTCCCTGTTACTCGGCGTCCCTCTGTACTCTGT 901

Db 138438 TCC-----AGCCTCCAGTCGACCCCTGTTATCATCCAGCGAGGTCTATGCACAGCG 138493

QY 902 ACTCTCATCTGAGCGCTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 961

Db 138494 ATTTCCCGCTCTGCT 138553

QY 962 TGCCCGCACTTGGCTG 977

Db 138554 GGCCCTCGCTCCCTG 138569

RESULT 9

MUSRARA2I

LOCUS MUSRARA2I 1330 bp DNA linear ROD 27-APR-1993

DEFINITION Mouse retinoic acid receptor alpha-2 isoform gene, 5' end.

ACCESSION M80781

VERSION M80781.1 GI:200654

KEYWORDS retinoic acid receptor-alpha-2-isoform.

SOURCE Mus musculus DNA.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 1330)

AUTHORS Leroy, P., Nakshatri, H. and Chambon, P.

TITLE MOUSE RETINOIC ACID RECEPTOR ALPHA-2 ISOFORM IS TRANSCRIBED FROM A PROMOTER THAT CONTAINS A RETINOIC ACID RESPONSE ELEMENT

JOURNAL Proc. Natl. Acad. Sci. U.S.A. (1991) In press

REFERENCE 2 (sites)

AUTHORS Leroy, P., Nakshatri, H. and Chambon, P.

TITLE Mouse retinoic acid receptor alpha 2 isoform is transcribed from a promoter that contains a retinoic acid response element

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (22), 10138-10142 (1991)

MEDLINE 92052226

PUBMED 1658797

FEATURES

Location/Qualifiers

1. 1330

source /organism="Mus musculus"

/db_xref="taxon:10090"

/cell_line="LMTK CELL"

/cell_type="FIBROBLAST"

prim_transcript 710..>1330

BASE COUNT 231 a 395 c 445 g 259 t

ORIGIN

Query Match 25.9%; Score 259; DB 10; Length 1330;

Best Local Similarity 64.8%; Pred. No. 7.7e-43;
Matches 633; Conservative 0; Mismatches 250; Indels 94; Gaps 13;

QY 34 CTTTTATATATGGGGTAATAGGATACCC-CCTCCTCCAGGGATATCCCTCTCTTTCTAGG 92
Db 64 CTTTTAGAGAGAGATTAATAGGATACCTCGTCTCTCTGGGTATCTCTTTCTAGG 123

QY 93 GACCTACCAAGCTAGGCTTTCTTCCAGTGAACGTCATCCGAGGGCTTCTAGGATG 152
Db 124 GACCTACCTAACCAAGCTTTCTTCCACTCAACAGCTGCTCTAG-----GGAGG 174

QY 153 AAGTAGTCCACTGAAGCACCAGCTCTCTCTTTATCTCTCCAGAGCTGACAGTGCAC 212
Db 175 AAGTGCTTCAACTTAAGTC-----TCCTTTTATCTGCTCTGGGT----- 214

QY 213 CAGGGCGCGGTACTGTTTCCACAGTAGGACACCTTGGCGGGGCTTCTCGCCGGA 272
Db 215 -----GGCCCATCTCTGTTCCGGCTGGGACACCTTGGCGGAGCTTCTCGCCGGA 270

QY 273 AGCAGCAGAGCTGGGAGAGGCGCCCTCTGCTGTGTGTGTGTCACACAGC----- 326
Db 271 AGCAGCAGAGCTGGGAGAGAGCTCCCC--GTTGTGTGTGTGCTTAACAGCTCAGTT 327

QY 327 -----ACCCGCGCTGCGCGCTGGGTTCGGGGCGCGGAGTCAACATGATGTACAC 378
Db 328 CTGCGCTGCTGCTTATTTCGGCGGGTGAGCCGCGAGCTCCACATGATGTACAC 387

QY 379 GACATGACACACCGCG--TGTCTCATTCGACACAGGTCGAGCTGCAACATGTACAC 437
Db 388 GACAGTGACACAGCTGGATGTCTCATTCGACACAGCGCCCAAGCTGCAACATGTACAC 447

QY 438 CCGGGGTGCCAAA-----CACTTGGCCCGCGCGAGCCGCGGCTACGCGCTC 483
Db 448 CCGGGACACCCAGCATGCTGCTGTTCTAAGCTGGGCGACCTGCAATTTGGGCC 507

QY 484 CTGCGCGCTCTCCGCGTCTCCGGGAGAGTGGCCCGGTTTCGCGCGGCGAGGGGCTGG 543
Db 508 CTGTGCTACTGCGCGTGTCTCGAGATTTGGGCGCGGACAGGTCGAGATGGGGCCAG 567

QY 544 CGGGCAGACCCCGCGCGCTGGCGAGCGGTGATGTACCGGACGCGGTGGTGGGT 603
Db 568 TGGCGAGCCCGCGAGCGGGTGGCGAGCTGGTGTGTACGGGCGAGCGGTGGTGGGT 627

QY 604 CACTCGAGGTAGGCGCCCGCAGCGAGTTCAGCGAGAGTTCAGCGCATTTGATTAGG 663
Db 628 CAGTCGAGGTAGGCGCCCGCAGCGAGTTCAGCAGAGTTCAGCGCAATTGCATTAGG 687

QY 664 CAATGAGCGCGGCC--TGGTGGGGGTGTGTTAAGGGAGGACACCGGACACCC 721
Db 688 CAATGAGCGCGGGCTTTGGGGGGGGGGTGTGTTAGGGGAGGACACCGGACACCC 747

QY 722 CCTCTTCCCGCCCGCCACACCTCTCCACACGCTTCTCGCGCGAGGACTGACCA- 780
Db 748 C-----ACCTCCCGCGCTGCTATCTCTGCGAGGGGATTGACCAAG 788

QY 781 AACCTTTGGGAGCTGGGAGCGGAACTGGTACAAGGGAGGAGACCGCGCCCTCTCTCC 840
Db 789 AAGCACTAGGAGGCTGGGAGCGGAACTGG--GATGGGAGAGATGCCCTCTACTCTCG 846

QY 841 GTCTTGTCCCTCTGCGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db 847 GTCC----AGCCCTCCAGTGCACCTCTGTTTCCATACCCAGCCAGGTTCTATGACAGCG 902

QY 901 TACTCTCTCATCTGAGGCTTTCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db 903 CATTTCCCGCTCTGTCT 962

QY 961 CTGCCCCCACTTGCCTG 977
Db 963 TGGCCCTCTGCTTCCCTG 979

RESULT 10

AX344791/c

LOCUS AX344791 13508 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 216 from Patent WO200927.
ACCESSION AX344791
VERSION AX344791.1 GI:18492677
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with development genes
JOURNAL Patent: WO 020927-A 216 03-JAN-2002;
EpiGenomics AG (DE)
FEATURES Location/Qualifiers
source 1..13508
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3338 a 286 c 3789 g 6094 t 1 others
ORIGIN

Query Match 24.3%; Score 242.8; DB 6; Length 13508;
Best Local Similarity 80.3%; Pred. No. 1.2e-39;
Matches ,297; Conservative 0; Mismatches 72; Indels 1; Gaps 1;

QY 631 AGTTCAGCGAGATTACGCCGCTTGCATTAGGCAATAGCGCCGCTGGTGGGGT 690
Db 665 AATTCACGAAATTCACCGCATTTACATTAACAATAAACCAGCCTAAATAAAAT 6606

QY 691 GTGTGTTAAGGGAGACACGGGACACCCCTCTTCCCGCCGACACCTCTCTCCA 750
Db 6605 ATATATTAATAAATAAACAACACGAAACACCCCTCTTCCCGCCGACACCTCTCTCCA 6546

QY 751 CCACGGCTTCGCTCGCCGAGGAGTACCAAACTTGGGGGAGCTGGGAGCGGAACTG 810
Db 6545 CCACGACTTCGCTCGACCAAAACTTAACCAAACTTAAAAAACTTAAAAACCGAACTA 6486

QY 811 GTACAGGGGAGGAGCGCGCCCTCTTCCGTCCTTCTTCCCGCTCTCGCAGCCCTCTCTC 870
Db 6485 ATACAA-AAAAAAGCGCGCCCTCTTCCGTCCTTATCCCTCGCAACCCCTCTCTC 6427

QY 871 CCTTACTCGCGCTCCCTCTCTGTACTCTGTACTCTCTCTCTCTCTCTCTCTCTCTCTCT 930
Db 6426 CCTTACTCGAGTCCCTCTCTATCTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6367

QY 931 TGTCT 990
Db 6366 TACTTCT 6307

QY 991 TCTCCCTCTC 1000
Db 6306 TCTCCCTCTC 6297

RESULT 11

LOCUS AK098172 2008 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ40853 fis, clone TRACH2015486, highly similar
to RETINOIC ACID RECEPTOR ALPHA.
ACCESSION AK098172
VERSION AK098172.1 GI:21758129
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens trachea cDNA to mRNA, clone_lib:TRACH2
clone:TRACH2015486.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Suzuki, O., Sasaki, N., Aotsuka, S., Shoji, T., Ichihara, T.,
Shiohata, N., Matsumoto, K., Hirano, M., Sano, S., Nomura, R.,
Yoshikawa, Y., Matsumura, Y., Moriya, S., Chiba, E., Momiyama, H.,
Onogawa, S., Kaeriyama, S., Satoh, N., Matsunawa, H., Takahashi, E.,

Kataoka, R., Kuga, N., Kuroda, A., Satoh, I., Kamata, K., Takami, S., Terashima, Y., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Watsushima, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagabari, K., Masuho, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project

TITLE
Unpublished
JOURNAL
2 (bases 1 to 2008)
REFERENCE
Isogai, T. and Yamamoto, J.
AUTHORS
Direct Submission
TITLE
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
JOURNAL
kazusa-kamata, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology (RAB); cDNA library construction; Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing; HRI and RAB; annotation: HRI and RAB.

FEATURES
source
Location/Qualifiers
1..2008
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TRACH2015486"
/tissue_type="trachea"
/clone_lib="TRACH2"
/note="cloning vector: pME18SFL3"

BASE COUNT
ORIGIN
405 a 669 c 576 g 358 t

Query Match 22.8%; Score 228; DB 9: Length 2008;
Best Local Similarity 100.0%; Pred. No. 1.5e-36;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 ACATGATGTCACAGCAATGACAAAGCGGTGCTCATTCGACACAGCGTCCGAGCTG 425
|||||
Db 1 ACATGATGTCACAGCAATGACAAAGCGGTGCTCATTCGACACAGCGTCCGAGCTG 60
|||||

QY 426 CACAATGTCACACCGGGTGCCAAACACTTGGCCCCGCGACCGCCCTACGCTCTCT 485
|||||
Db 61 CACAATGTCACACCGGGTGCCAAACACTTGGCCCCGCGACCGCCCTACGCTCTCT 120
|||||

QY 486 GCCCGCTCTCCGCTCCGGGGAGGTGGCGGTTCGGCGGCGAGGGGCTGGCG 545
|||||
Db 121 GCCCGCTCTCCGCTCCGGGGAGGTGGCGGTTCGGCGGCGAGGGGCTGGCG 180
|||||

QY 546 GCGGAGCGCGCGGCTGGGAGCGGTGATGTCACGGGACGG 593
|||||
Db 181 GCGGAGCGCGCGGCTGGGAGCGGTGATGTCACGGGACGG 228
|||||

RESULT 12
AX344790
LOCUS
AX344790
DEFINITION
Sequence 215 from Patent WO0200927.
ACCESSION
AX344790
VERSION
AX344790.1 GI:18492676
KEYWORDS
synthetic construct.
SOURCE
synthetic construct
artificial sequences.
ORGANISM
1
REFERENCE
Olek, A., Piepenbrock, C. and Berlin, K.
AUTHORS
Diagnosis of diseases associated with development genes
TITLE
Patent: WO 0200927-A 215 03-JAN-2002;
JOURNAL
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
1..13508

/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT
ORIGIN
2350 a 286 c 4030 g 6841 t

Query Match 14.4%; Score 143.8; DB 6: Length 13508;
Best Local Similarity 53.2%; Pred. No. 1.8e-19;
Matches 237; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

QY 626 AGGCGAGTTACGAGAGTTCAGCGCATTCATTAGGCAATGAGCCCGCGCTGGGTG 685
|||||
Db 6839 AGGGTAGTTACGAGAGTTCAGCGCATTCATTAGGCAATGAGCTTCGGTTGGGTG 6898
|||||

QY 686 GGGGTGTGTAAAGGGGAGGACCGGGGACCCCTCTTCCCGCCGCCACACCTC 745
|||||
Db 6899 GGGGTGTGTAAAGGGGAGGATATCGGGATTATTTTTTTTTTTTCGTTTATTATT 6958
|||||

QY 746 CTCACACCGGTTCTCGCTCGGCGGAGTGTACCAACCTTGGGGAGCTGGAGCGCG 805
|||||
Db 6959 TTTTATTACGGTTCTCGGTTAGGATGTAAATTTGGGGAGTTGGGAGTCGG 7018
|||||

QY 806 AACTGGTACAAGGGGAGGACGCGCCCTCTTCCCTCTTCTCCCTCGCACGCCCTC 865
|||||
Db 7019 AATTGGTATAA-GGGAGGACGTTCTGTTTTTTTTCGTTTTTTCGTAGTTTTT 7077
|||||

QY 866 CTCTCCCTTACTCGCGCTCTCTGTACTCTGTGTACTCTCTCATCTGAGCCTTTCCC 925
|||||
Db 7078 TTTTTCGTTATTCGCGCTTTTTCGTTATTTGTTGTTATTTTATTTGGAGTTTTTT 7137
|||||

QY 926 CTTCCTGCTTCTCTCTCTCCCTCCCGCTCCAGGCGTCCCGACCTGCTGTCACATG 985
|||||
Db 7138 TTTTTCGTTTTTTTTTTTTTTTTTTTATTTAGGTGTTTTTATTTGTTTATATG 7197
|||||

QY 986 CGCGCTCTCCCTCTC 1000
|||||

Db 7198 TCGTTTTTTTTTTTC 7212
|||||

RESULT 13
AC111746/c
LOCUS
AC111746
DEFINITION
Rattus norvegicus clone CH230-15011, *** SEQUENCING IN PROGRESS
***, 72 unordered pieces.
ACCESSION
AC111746
VERSION
AC111746.2 GI:21736893
KEYWORDS
HTG; HTGS, PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 162393)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayelle, M., Banks, T.,
Barbaria, J., Benton, J., Blincoe, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brivea, M., Brown, E., Brown, N.P., Bryant, N.P.,
Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davilla, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,

Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lounsbury, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., McLeod, M.P., Meador, M., Mei, G., Metzker, E., Miner, G., Minner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scheraga, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmari, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE

Direct Submission

REFERENCE

2 (bases 1 to 162393)

AUTHORS

Worley, K.C.

TITLE

Direct Submission

REFERENCE

Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

AUTHORS

Worley, K.C.

TITLE

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 12, 2002 this sequence version replaced gi:18701588.
----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GOAF

Center clone name: CH230-15011

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 96609 bases at least Q40

Consensus quality: 101360 bases at least Q30

Consensus quality: 105362 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 72 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1242: contig of 1242 bp in length
* 1243 1342: gap of unknown length
* 1343 2427: contig of 1085 bp in length
* 2428 2527: gap of unknown length
* 2528 3635: contig of 1108 bp in length
* 3636 3735: gap of unknown length
* 3736 4821: contig of 1086 bp in length
* 4822 4921: gap of unknown length
* 4922 6640: contig of 1719 bp in length
* 6641 6740: gap of unknown length
* 6741 8216: contig of 1476 bp in length
* 8217 8316: gap of unknown length

8317 9346: contig of 1030 bp in length
* 9347 9446: gap of unknown length
* 9447 10691: contig of 1245 bp in length
* 10692 10791: gap of unknown length
* 10792 11851: contig of 1060 bp in length
* 11852 11951: gap of unknown length
* 11952 13509: contig of 1558 bp in length
* 13510 13609: gap of unknown length
* 13610 14869: contig of 1260 bp in length
* 14870 14969: gap of unknown length
* 14970 16275: contig of 1306 bp in length
* 16276 16375: gap of unknown length
* 16376 17441: contig of 1065 bp in length
* 17442 17540: gap of unknown length
* 17541 18956: contig of 1416 bp in length
* 18957 19056: gap of unknown length
* 19057 20611: contig of 1005 bp in length
* 20612 20662 20161: gap of unknown length
* 20663 21232: contig of 1071 bp in length
* 21233 21332: gap of unknown length
* 21333 22378: contig of 1046 bp in length
* 22379 22478: gap of unknown length
* 22479 23629: contig of 1151 bp in length
* 23630 23729: gap of unknown length
* 23730 24993: contig of 1264 bp in length
* 24994 25094 25093: gap of unknown length
* 25095 26648: contig of 1555 bp in length
* 26649 26748: gap of unknown length
* 26749 28383: contig of 1635 bp in length
* 28384 28483: gap of unknown length
* 28484 28583: gap of unknown length
* 28584 29659: contig of 1175 bp in length
* 29660 31548: contig of 1790 bp in length
* 31549 31649: contig of 1267 bp in length
* 31650 32915: contig of 1267 bp in length
* 32916 33015: gap of unknown length
* 33016 34132: contig of 1117 bp in length
* 34133 34232: gap of unknown length
* 34233 35376: contig of 1144 bp in length
* 35377 35476: gap of unknown length
* 35477 36835: contig of 1359 bp in length
* 36836 36935: gap of unknown length
* 36936 38119: contig of 1184 bp in length
* 38120 38219: gap of unknown length
* 38220 40048: contig of 1829 bp in length
* 40049 40148: gap of unknown length
* 40149 41648: contig of 1500 bp in length
* 41649 41748: gap of unknown length
* 41749 44023: contig of 2275 bp in length
* 44024 44123: gap of unknown length
* 44124 45297: contig of 1174 bp in length
* 45298 45397: gap of unknown length
* 45398 46533: contig of 1136 bp in length
* 46534 46633: gap of unknown length
* 46634 48451: contig of 1818 bp in length
* 48452 48551: gap of unknown length
* 48552 49889: contig of 1338 bp in length
* 49890 49989: gap of unknown length
* 49990 52066: contig of 2077 bp in length
* 52067 52166: gap of unknown length
* 52167 53724: contig of 1558 bp in length
* 53725 53824: gap of unknown length
* 53825 55350: contig of 1526 bp in length
* 55351 57835: contig of 2385 bp in length
* 57836 57935: gap of unknown length
* 57936 59470: contig of 1535 bp in length
* 59471 59570: gap of unknown length
* 59571 61049: contig of 1479 bp in length
* 61050 61149: gap of unknown length
* 61150 62619: contig of 1470 bp in length
* 62620 65125: contig of 2406 bp in length
* 65126 65125: contig of 2406 bp in length

```
* 65126 65225: gap of unknown length
* 65226 65227: contig of 1202 bp in length
* 65428 65527: gap of unknown length
* 66528 68713: contig of 2186 bp in length
* 68814 68813: gap of unknown length
* 71416: contig of 2603 bp in length
* 71516: gap of unknown length
* 71517 73660: contig of 2144 bp in length
* 73661 73760: gap of unknown length
* 73761 75002: contig of 1742 bp in length
* 75003 77965: gap of unknown length
* 77966 78065: contig of 2363 bp in length
* 78066 79509: contig of 1444 bp in length
* 79510 79608: gap of unknown length
* 79610 81649: contig of 2040 bp in length
* 81650 81749: gap of unknown length
* 81750 83575: contig of 1826 bp in length
* 83576 83675: gap of unknown length
* 83676 86277: contig of 2602 bp in length
* 86278 86377: gap of unknown length
* 86378 88701: contig of 2324 bp in length

Query Match
Best Local Similarity 12.7%; Score 127; DB 2; Length 162393;
Matches 254; Conservative 0; Mismatches 85; Indels 26; Gaps 5;

QY 562 GGGTGGCGAGCGGTGATGTACAGCGGAGCGGTGGGTGCTGCTGAGTGGAGCGC 621
Db 142398 GGGTGGCGAGCGGTGATGTACAGCGGAGCGGTGGGTGCTGCTGAGTGGAGCGC 142339

QY 622 CGCCAGCGAGCTTACAGCAGATTCAGCGGATTCATAGCAATAGAGCCCGCCCT- 680
Db 142338 CGTCAGCGAGTTCAGCAAGAGTTTCAGTCGATTCATAGCAATAGAGTGGGGGTG 142279

QY 681 -GGGTGGGGGTGTGTGTTAAGGGAGGAGACACCGGAGCACCCTCTTCCCGCCCGAC 739
Db 142278 GGGTGGGGGGCGTGTTCAGGGAGAGACCCAGCAGCACCCTCTTCCCGCCCGAC 142236

QY 740 CACCTCTTCCAGCGGTTCGCTCGGCGAGGAGTACGAC- AAACCTTGGGGAGCGTGG 798
Db 142235 -ACCTCTCCCGCGTGTTCGCGCAGCGGAGTTCACCAAGAGTACGAGGAGCGCG 142178

QY 799 GAGCGGAGCTGTTACAGGAGGAGGAGCGCGCCCTCTTCCGCTGCTGCTCCCTCGAG 858
Db 142177 GTGCCAGAACTGG--GATGGGAGAGTCTCTCTCTGCTGGTCTCT--GCCTCCGGTG 142122

QY 859 CCCCTCTCTCCCTGTACTGCGGCTCCCTGCTGTACTCTGTACTCTCTCATCTGGAGCC 918
Db 142121 CACCTGTATCATACCGACCTGGTTCATGACAGCGCTTTCCCGCCTCTGCTCCC 142062

QY 919 TTTC 923
Db 142061 TCTCC 142057

RESULT 14
PM12A12G/c 972 bp DNA linear STS 09-MAR-2002
LOCUS Penicillium marneffei STS, clone pm12a12.g, sequence tagged site.
DEFINITION
ACCESSION AL684288
VERSION AL684288.1 GI:19338015
KEYWORDS STS.
SOURCE Penicillium marneffei.
ORGANISM Penicillium marneffei.
REFERENCE 1
AUTHORS Yuen K.Y., Pascal, G., Wong, S., Glaser, P., Woo, P., Kunst, P.,
TITLE Cheung, E., Medigue, C. and Danchin, A.
JOURNAL Exploring the Penicillium marneffei genome
Unpublished
REFERENCE 2 (bases 1 to 972)
```

```
AUTHORS Danchin, A. and Pascal, G.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES
Location/Qualifiers
source
1..972
/organism="Penicillium marneffei"
/db_xref="taxon:37727"
/clone="pm12a12.g"
BASE COUNT 53 a 334 c 428 g 20 t 137 others
ORIGIN
Query Match
Best Local Similarity 7.4%; Score 73.8; DB 11; Length 972;
Matches 219; Conservative 0; Mismatches 291; Indels 0; Gaps 0;

QY 489 GCCGCTCTCCGCTCTCCGGGGAGGTGGCCCGTTCGGCGGAGGCGGCTGGCGGGC 548
Db 835 GCGGGGGTGTGGCCCGGGGGGCGCCCGGGGGGGCGGGGGGGCGGGGGGGGG 776

QY 549 GAGCCCCCGGGGGGCTGGCGAGCGGTGATGTACAGGCGAGCGGTGGGTCACTC 608
Db 775 GGGGGGGGGTGGGGGGGGGGGGGGTCCCGCCGCGGGGGGGGGGGGGCGCGC 716

QY 609 GGAGGTGAGCGCGCCGAGCGAGTTTCAGCGAGTTCAGCCGATTCATTAGGCAAT 668
Db 715 NCGGGCGGTGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 656

QY 669 GAGGCGCGCGCTGGGTGGGGGTGTGTTAAGGGAGGAGACCGGGAGCACCCTCTT 728
Db 655 GGGGGCGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 596

QY 729 CCGCGCGCCACACACCTCTCCACCGGTTCGCTCGGCGAGGAGTACCAACCTTGG 788
Db 595 GGTGCGCGGGGGGGTCCNCCGCTCGCGCCCCCCCCCCCCCCCCCCCCCG 536

QY 789 GSGAGCTGGAGCGGAACTGTGTACAGGGAGGAGCGCGCCCTCTTCCGCTTGT 848
Db 535 GGTTCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 476

QY 849 CCGCTGCGACCGCCCTCTCTCCCTGCTACTCGGCGTCCCTCTGTACTGTCTCTC 908
Db 475 CCGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 416

QY 909 AATCGAGCGCTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 968
Db 415 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 356

QY 969 ACTTGCTGTGCACATGCGCGCTCTCCCTC 998
Db 355 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326

RESULT 15
I66494
LOCUS I66494 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
source
1..7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
```

```

Query Match      7.3%; Score 72.6; DB 6; Length 7218;
Best Local Similarity 11.5%; Pred. No. 6.2e-05;
Matches 27; Conservative 142; Mismatches 66; Indels 0; Gaps 0;

QY 766 GCCAGGGACTGACCAACCTTGGGGGAGCCCTGGGAGCCGGAACCTGGTACAGGGGAGGAC 825
    |||| || || || |||| || || || || || || || || || || || || || ||
Db 1010 GCCATACGCTCACAGAAATAATCCGAGCTTGGCTGCGAGTTCGAGGAGCTTGCATYYY 1069

QY 826 GCCCGCCCTCTTCGTCCTTGTCCCTCGCAGCCCTCCTCTCTCCCTGTACTCGCGTC 885
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1070 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1129

QY 886 CCTGTGACTGTGTACTCTCATCTGGAGCCTTCCCTTCTCTCTCTCTCTCTCTCTCTC 945
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1130 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1189

QY 946 CTCCCTTCCAGGCTGCCCTGTCACATGCCGCTCTCCCTCTC 1000
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1190 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1244
  
```

Search completed: March 30, 2003, 12:31:19
 Job time : 2702.41 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: March 30, 2003, 07:57:38 ; Search time 196.103 Seconds
(without alignments)
11483.757 Million cell updates/sec

Title: US-09-691-220-3_COPY_1_1000

Perfect score: 1000

Sequence: 1 gtccttggttagcatgtaca.....acatgcgccttcctctc 1000

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Result No.	Score	Query Match	Length	ID	Description
1	1000	100.0	20512	24	AAAL38339
c	242.8	24.3	13508	24	ABN80199
3	143.8	14.4	13508	24	ABN80198
4	59.2	5.9	53522	24	RAD30228
5	59.2	5.9	53526	19	AAT94101
6	59.2	5.9	53577	17	AAT18551
7	59.2	5.9	53577	19	AAT94108
c	56.8	5.7	114955	20	AAAX53491
9	54.4	5.4	114955	20	AAAX53491

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

10	54	5.4	13862	22	ABA08208	Human ovarian and
11	54	5.4	13862	22	AAU02789	Human reproductive
12	54	5.4	13862	22	AAU07516	Human reproductive
c	51.6	5.2	4897	11	AAQ03259	Pseudorabies virus
14	51.4	5.1	600	24	ABQ52496	Oligonucleotide fo
15	51.4	5.1	600	24	ABQ52497	Oligonucleotide fo
16	50.6	5.1	2188	20	AAZ77506	Human ovarian tumo
17	50.4	5.0	154746	24	AAZ25519	Human herpesvirus
c	50.4	5.0	154746	24	AAZ25519	Human herpesvirus
19	49.8	5.0	1337	20	AAZ17263	Human gene express
c	49.6	5.0	320	21	AAZ38186	Primer used in the
21	49.4	4.9	2250	22	AAH17884	Human cDNA sequenc
c	49.2	4.9	318	21	AAH38184	Primer used in the
23	49	4.9	1000	21	AAQ02484	Human colon cancer
c	48.8	4.9	1127	21	AAA02477	Human colon cancer
25	48.6	4.9	3978	21	AAA53918	Adenyl cyclase t
c	48.4	4.8	1327	24	ABQ68452	Listeria monocytog
27	48.4	4.8	3198	20	AAQ02974	Human IL-1ra BAC c
c	48.4	4.8	43058	24	ABN97455	Gene #3953 used to
28	48.4	4.8	43058	24	ABN97455	Lung cancer relate
29	48.4	4.8	43058	24	ABL64982	Lung cancer relate
c	48.2	4.8	43058	24	ABL65219	Oligonucleotide fo
31	48.2	4.8	693	24	ABQ32102	Oligonucleotide fo
c	48.2	4.8	693	24	ABQ32103	Oligonucleotide fo
33	48.2	4.8	105325	24	ABK94407	DNA encoding endot
c	48	4.8	320	21	AAA38183	Primer used in the
35	47.8	4.8	320	21	AAA38185	Primer used in the
c	47.2	4.7	1385	24	ABQ68566	Listeria monocytog
37	47.2	4.7	1385	24	ABQ70117	Listeria monocytog
c	46.4	4.6	1327	24	ABQ68452	Listeria monocytog
39	46.2	4.6	598	24	ABQ52556	Oligonucleotide fo
c	46.2	4.6	598	24	ABQ52557	Oligonucleotide fo
41	45.8	4.6	630	24	ABQ40868	Oligonucleotide fo
c	45.8	4.6	630	24	ABQ40869	Oligonucleotide fo
43	45.6	4.6	186	17	AAT35033	HSV-1 IR-L (positi
c	45	4.5	2561	22	AAH26500	Rabbit low density
c	45	4.5	7694	22	ABA07755	Human ovarian and

ALIGNMENTS

RESULT 1

AAAL38339
ID AAL38339 standard; DNA; 20512 BP.

XX AC AAL38339;

XX AC AAL38339;

DT 15-AUG-2002 (first entry)

XX Genomic DNA encoding the human nuclear hormone receptor protein.
DE DE Genomic DNA encoding the human nuclear hormone receptor protein.
XX KW Human nuclear hormone receptor; therapeutic agent; gene therapy;
KW Immune response; chromosome 17; single nucleotide polymorphism;
KW gene; ds.

XX Homo sapiens.

XX OS Homo sapiens.

XX Key

FT CDS

FT Location/Qualifiers

FT 1684...15208

FT /*tag= a

FT /product= "Human nuclear hormone receptor protein"

FT 1684...1846

FT /*tag= b

FT /number= 1

FT 1847...6808

FT /*tag= c

FT /number= 1

FT 6809...6957

FT /*tag= d

FT /number= 2

FT 6958...8277

FT /*tag= e

FT /number= 2

Db 601 GGTCACTCGAGGTGAGCGCGCGGAGGAGTTCAGCGAGAGTTCAGCGCATTCATT 660
Qy 661 AGGCAATGAGCGCGCGCTGGTGGGGTGTGTAAAGGGAGGACACCGGGACACC 720
Db 661 AGGCAATGAGCGCGCGCTGGTGGGGTGTGTAAAGGGAGGACACCGGGACACC 720
Qy 721 CCCCTCTTCCCGCCCGCCACCTCTCCACCGGCTTCGCTCGGCGAGGACTGACCA 780
Db 721 CCCCTCTTCCCGCCCGCCACCTCTCCACCGGCTTCGCTCGGCGAGGACTGACCA 780
Qy 781 AACCTTGGGAGGCTGGAGCGGCACTGTGTACAAAGGGAGGACCGCGCCCTTCC 840
Db 781 AACCTTGGGAGGCTGGAGCGGCACTGTGTACAAAGGGAGGACCGCGCCCTTCC 840
Qy 841 GTCCTTGTCCCTCGCAGCGCCCTCTCTCCCTGTACTCGGGTCCCTGTACTCTGTG 900
Db 841 GTCCTTGTCCCTCGCAGCGCCCTCTCTCCCTGTACTCGGGTCCCTGTACTCTGTG 900
Qy 901 TACTCTCATCTGGAGCGCTTTCCTCCCTTCCCTGTCTCTCTCTCTCTCTCTCTCT 960
Db 901 TACTCTCATCTGGAGCGCTTTCCTCCCTTCCCTGTCTCTCTCTCTCTCTCTCTCT 960
Qy 961 CTGCCCCACTTGCTGTCCACATGCGCGCTCTCCCTCTC 1000
Db 961 CTGCCCCACTTGCTGTCCACATGCGCGCTCTCCCTCTC 1000

RESULT 2

ABN80199/c
ID ABN80199 standard; DNA; 13508 BP.

AC ABN80199;

XX 15-JUL-2002 (first entry)

DE Human chemically modified disease associated gene SEQ ID NO 216.

XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.

OS Homo sapiens.
OS Synthetic.

XX WO200200927-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07536.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130908/17.

XX Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a
PT sequence of a segment of chemically pretreated DNA of genes associated
PT with development

XX Claim 1; SEQ ID NO 216; 27pp; English.

XX The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in
CC the specification such as ACCPN, ADFN, or AFDI and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with

CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, (iii) or their complements for
CC the amplification of the 350 sequences, (iii) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs).

CC Note: The sequence data for this patent did not form part of the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.

XX SQ Sequence 13508 BP; 3338 A; 286 C; 3789 G; 6094 T; 1 other;

Query Match 24.3%; Score 242.8; DB 24; Length 13508;
Best Local Similarity 80.3%; Pred No. 8.3e-45;
Matches 297; Conservative 0; Mismatches 72; Indels 1; Gaps 1;

Qy 631 AGTTCAGCGAGAGTTCAGCGCATTCATTAGGCAATAGAGCGCGCTGGTGGGGT 690
Db 6665 AATTCACGAAATTCACCGCATTCATTAAACAATAAAACCCGACCTAAATAAAAT 6606

Qy 691 GTGTGTAAAGGGAGGACACCGGACCCCTCTTCCCGCCCGCCACCTCTCTCCA 750
Db 6605 ATATATTAATAAAACACCGGACCCCTCTTCCCGCCCGCCACCTCTCTCCA 6546

Qy 751 CCACCGCTTCGCTCGCGGAGGACTGACCAAACTTGGGGAGCTTGGGAGCGGACTG 810
Db 6545 CCACGACTTCGCTCGCGGAGGACTGACCAAACTTGGGGAGCTTGGGAGCGGACTG 6486

Qy 811 GTACAGGGGAGGAGCGCGCGCTTCTTCCGCTCTTGTCCCGCTCGAGCGCCCTCTCTC 870
Db 6485 ATACAA-AAAAAAGCGCGCGCTTCTTCCGCTCTTGTCCCGCTCGAGCGCCCTCTCTC 6427

Qy 871 CCGTACTCGCGCTCCCTCTGTACTCTGTACTCTCTCTCTCTCTCTCTCTCTCTCT 930
Db 6436 CCGTACTCGCGCTCCCTCTGTACTCTGTACTCTCTCTCTCTCTCTCTCTCTCTCT 6367

Qy 931 TCGTCT 990
Db 6366 TACTTCT 6307

Qy 991 TCTCCCTCTC 1000
Db 6306 TCTCCCTCTC 6297

RESULT 3
ABN80198

ID ABN80198 standard; DNA; 13508 BP.

XX AC ABN80198;

XX 15-JUL-2002 (first entry)

XX Human chemically modified disease associated gene SEQ ID NO 215.

XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.

OS Homo sapiens.
OS Synthetic.

XX WO200200927-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07536.

XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIC-) EPIGENOMICS AG.
XX Olek A, Plepenbrock C, Berlin K;
PI WPI; 2002-130908/17.
XX
XX Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a
PT sequence of a segment of chemically pretreated DNA of genes associated
XX with development
XX
PS Claim 1; SEQ ID NO 215; 27pp; English.
XX
XX The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in
CC the specification such as ACCPN, ADPN, or AFDI and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with
CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currairino syndrome, diseases related to the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs).
CC Note: The sequence data for this patent did not form part of the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
XX Sequence 13508 BP; 2350 A; 286 C; 4030 G; 6841 T; 1 other;
SQ
Query Match 14.4%; Score 143.8; DB 24; Length 13508;
Best Local Similarity 63.2%; Pred. No. 1.1e-22;
Matches 237; Conservative 0; Mismatches 137; Indels 1; Gaps 1;
QY 626 AGCGAGTTCAGCGAGAGTTTCAGCGCATTCATTAGGCAATCAGCCCGCGCTGGTG 685
DB 6839 AGGTAGTTTACGAGAGTTTATGTCGTTATGATGTAATGAGTTCGGTTGGTG 6898
QY 686 GGGGTGTGTCTTAAGGGAGGACACCGGACACCCCTCTCCCGCCGCCACACCTC 745
DB 6899 GGGGTGTGTCTTAAGGGAGGATATCGGATATTTTTTTTTTTCGTTTATATTTT 6958
QY 746 CTCACACAGCGTTCGTCGGCGAGGACGACCAACCTTGGGGGACCTGGAGCGCG 805
DB 6959 TTTTATTACGGTTTCGTCGTTAGGATGATTAATTTTGGGGAGTTGGGAGTCGG 7018
QY 806 AACTGTACAAAGGGAGGACCGCCCTCTCCGTCCTTCCTCCGAGCCCTC 865
DB 7019 AATTGGTATAA-GGGAGGAGCTGCTTTTTTTTCGTTTTTTCGTTAGTTT 7077
QY 866 CTCTCCCTGTACTCGGCTCCCTCTGTACTCTCTCTCTCTCTCTCTCTCTCTCTCT 925
DB 7078 TTTTATTTTGTATTCGGCTTTTGTATTTTGTATTTTATTTTATTTTATTTTATTT 7137
QY 926 CTTTCGCT 985
DB 7138 TTTTATTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATG 7197
QY 986 CGGCT 1000
DB 7198 TCGTTTTTTTTTTTTC 7212

RESULT 4

AAD30228
ID AAD30228 standard; DNA; 53522 BP.
XX
XX AAD30228;
XX
XX 17-MAY-2002 (first entry)
XX Human PKD1 gene.
XX
XX Human; PKD1 gene; autosomal dominant polycystic kidney disease; ADPKD;
KW acquired cystic disease; transgenic animal; chromosome 16; ds.
KW
XX Homo sapiens.
OS
XX WO200206529-A2.
PN
XX 24-JAN-2002.
PD
XX 13-JUL-2001; 2001WO-US22035.
PF
XX 13-JUL-2000; 2000US-218261P.
PR
XX 13-APR-2001; 2001US-283691P.
PR
XX (UJJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PA
XX Germino GG, Watnick TJ, Phakdeekitcharoen B;
PI WPI; 2002-179805/23.
XX
XX Novel primer for diagnosing polycystic kidney disease-associated
PT disorder, comprises regions having sequence that selectively hybridizes
PT to polycystic kidney disease gene sequence
XX
XX Claim 20; Page 127-156; 192pp; English.
XX
XX The present invention relates to compositions and methods useful for the
CC identification and detection of polycystic kidney disease (PKD1) gene
CC mutations. The invention also relates to primers comprising a 5' region
CC having a sequence that selectively hybridizes to a PKD1 gene sequence
CC and optionally, to a PKD1 homologue sequence and an adjacent 3' region
CC having a sequence that selectively hybridizes to a PKD1 gene sequence
CC and not to a PKD1 homologue sequence. Primer pairs of the invention are
CC useful for detecting the presence or absence of a mutation in a PKD1
CC polynucleotide in a sample, for identifying a subject at risk for a
CC PKD1-associated disorder such as autosomal dominant polycystic kidney
CC disease (ADPKD) or acquired cystic disease and for diagnosing a PKD1-
CC associated disorder in a subject. They are useful for selectively
CC amplifying a region of a PKD1 gene. PKD1 DNA fragments are useful
CC detecting the presence of a mutant PKD1 polynucleotide in a sample,
CC as a probe for an amplification reaction, in hybridization or
CC amplification assays of biological samples to detect abnormalities
CC of PKD1 expression and for engineering transgenic animals. The present
CC sequence is human PKD1 gene located on chromosome 16.
XX
XX Sequence 53522 BP; 8486 A; 17664 C; 15766 G; 11606 T; 0 other;
SQ
Query Match 5.9%; Score 59.2; DB 24; Length 53522;
Best Local Similarity 58.5%; Pred. No. 0.0011;
Matches 103; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 825 CGCCCGCCCTCTTCGCTGTGTCCTCGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 884
DB 35409 CCCCCTCTCTCCATCCCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 35468
QY 885 CCCT 944
DB 35469 CCCT 35528
QY 945 CCTCCCTCTCTCCAGGCTGCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1000
DB 35529 CCTCCCT 35584

Db	35528	CCCTCCCTCCCTTCCTCCCTCCCTTCCTCCCTCCCTTCCTCCCTCCCTTCCTCCCTTCCTC	35983
RESULT 8			
AXX53491/C			
ID	AXX53491	standard: DNA; 114955 BP.	
XX	AXX53491;		
XX	05-JUL-1999	(first entry)	
XX	Human adenosine A1 receptor antisense oligonucleotide fragment.		
DE	XX	Antisense oligonucleotide; multiple target; antisense treatment;	
KW	XX	impaired respiration; inflammation; lung disease;	
KW	XX	pulmonary vasoconstriction; inflammation; allergic rhinitis;	
KW	XX	acute asthma; allergy; asthma; impeded respiration;	
KW	XX	respiratory distress syndrome; pain; cystic fibrosis;	
KW	XX	pulmonary hypertension; pulmonary vasoconstriction; emphysema;	
KW	XX	chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;	
KW	XX	colon cancer; breast cancer; lung cancer; pancreatic cancer;	
KW	XX	hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastases	
KW	XX	prostate cancer; ss.	
OS	XX	Synthetic.	
PN	XX	WO9913886-A1.	
PN	XX	25-MAR-1999.	
PD	XX	17-SEP-1998; 98WO-US19419.	
PF	XX	09-JUN-1998; 98US-0093972.	
PR	XX	17-SEP-1997; 97US-0059160.	
PR	XX	(UYEC-) UNIV EAST CAROLINA.	
PA	XX	Nyce JW;	
PI	XX	WPI; 1999-229400/19.	
PI	XX	New antisense oligonucleotides used in treatment of, e.g. pulmonary	
DR	XX	vasoconstriction	
PT	XX	Disclosure; Page 37; 120pp; English.	
PT	XX	The specification describes antisense oligonucleotides (AXX52869-X55528)	
CC	XX	directed against at least 2 mRNAs selected from target genes, coding	
CC	XX	non-coding regions of RNAs corresponding to target genes, gene	
CC	XX	initiation codons, genomic flanking regions, intron-exon borders, the	
CC	XX	5'-end, the 3'-end and the juxta-section between coding and non-coding	
CC	XX	regions and all segments of RNAs encoding proteins associated with one	
CC	XX	or more diseases, conditions or mixtures. The antisense oligonucleot	
CC	XX	may be derived from sequences AXX5272-74. These multiple target	
CC	XX	oligonucleotides (specifically AXX55180-271) can be used for the	
CC	XX	antisense treatment of diseases and conditions. Typical diseases and	
CC	XX	conditions are those associated with impaired respiration and	
CC	XX	inflammation, including lung diseases, pulmonary vasoconstriction,	
CC	XX	inflammation, allergic rhinitis, acute asthma, allergies, asthma, im	
CC	XX	respiration, respiratory distress syndrome, pain, cystic fibrosis,	
CC	XX	pulmonary hypertension, pulmonary vasoconstriction, emphysema, chron	
CC	XX	obstructive pulmonary disease (COPD), and cancers such as leukemias,	
CC	XX	lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,	
CC	XX	pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,	
CC	XX	hepatic metastases, as well as all types of cancers which may metast	
CC	XX	or have metastasized to the lungs, including breast and prostate can	
XX	XX	Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;	
XSQ			
Query Match		5.7%; Score 56.8; DB 20; Length 114955;	
Best Local Similarity		31.5%; Pred. No. 0.0043;	
Matches 220; Conservative		77; Mismatches 397; Indels 4;	

Db	104439	CCGCCCSNNNDNNCCGCGGCCGCGCGCGCCGCGCCSNNNDNNCCGCGGCCGCGG	104498
Qy	671	GGCCCGCCCTGGTGGGGTGTGTTAAGGGGAGGACACGGGACACCCCTCTTCC	730
Db	104499	CGCGCCCGCCSNNNDNNCCGCGGCCGCGCGCCGCGCCGCGSNNNDNNCCGCGGCCB	104558
Qy	731	CGCGCCACCACTCTCCACCAAGGCTTCGTCGGCGAGGACACTGACCAAACTTGGG	790
Db	104559	GGCGCGCCCGCCSNNNDNNCCGCGGCCGCGCGCCGCGCCSNNNDNNCCGCGGCCB	104618
Qy	791	GAGCCTGGGAGCGGAAGTGTACAAAGGGAGGAGCGCGCCCTCTTCCTGCTTGTCC	850
Db	104619	GGCGCGCCGCSNNNDNNCCGCGGCCGCGCGCCGCGCCSNNNDNNCCGCGGCCGCGC	104678
Qy	851	CTCGCAGCCCTCTCTCTCCCTGTACTCGCGCTCTCTGTACTCTGTGTACTCTCAT	910
Db	104679	GCCSNNNDNNCCGCGGCCGCGCGCCGCGCCSNNNDNNCCGCGGCCGCGCCSNNNDNN	104738
Qy	911	CTGAGCCCTTCCCT	960
Db	104739	CGBGCGCGCGCCSNNNDNNCCGCGGCCGCGCGCCGCGCCGCGCCSNNNDNNCCGCG	104788
RESULT 10			
ABA08208			
ID	ABA08208	standard; DNA; 13862 BP.	
XX	XX		
AC	ABA08208;		
XX	XX		
DT	11-JAN-2002	(first entry)	
XX	XX		
DE	Human ovarian and breast cancer associated polynucleotide SEQ ID NO 1003.		
XX	XX		
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;		
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antifungal;		
KW	vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;		
KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;		
KW	neurological disease; infection; human; secreted protein; ds.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	WO200155325-A2.		
XX	XX		
PD	02-AUG-2001.		
XX	XX		
PF	17-JAN-2001; 2001WO-US01345.		
XX	XX		
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225271.		
PR	14-AUG-2000; 2000US-0225272.		
PR	14-AUG-2000; 2000US-0225273.		
PR	14-AUG-2000; 2000US-0225274.		
PR	14-AUG-2000; 2000US-0225275.		
PR	14-AUG-2000; 2000US-0225276.		
PR	14-AUG-2000; 2000US-0225277.		
PR	14-AUG-2000; 2000US-0225278.		
PR	14-AUG-2000; 2000US-0225279.		
PR	14-AUG-2000; 2000US-0225280.		
PR	14-AUG-2000; 2000US-0225281.		
PR	14-AUG-2000; 2000US-0225282.		
PR	14-AUG-2000; 2000US-0225283.		
PR	14-AUG-2000; 2000US-0225284.		
PR	14-AUG-2000; 2000US-0225285.		
PR	14-AUG-2000; 2000US-0225286.		
PR	14-AUG-2000; 2000US-0225287.		
PR	14-AUG-2000; 2000US-0225288.		
PR	14-AUG-2000; 2000US-0225289.		
PR	14-AUG-2000; 2000US-0225290.		
PR	14-AUG-2000; 2000US-0225291.		
PR	14-AUG-2000; 2000US-0225292.		
PR	14-AUG-2000; 2000US-0225293.		
PR	14-AUG-2000; 2000US-0225294.		
PR	14-AUG-2000; 2000US-0225295.		
PR	14-AUG-2		